ABSTRACT

Title of Document: IDENTIFYING AND TRACKING MARINE PROTEIN AND ITS IMPORTANCE IN THE NITROGEN CYCLE USING PROTEOMICS

Elisha Kelly Moore, Ph.D., 2011

Directed By: Professor H. Rodger Harvey, Marine Estuarine and Environmental Science

Protein comprises the largest compartment of organic nitrogen in the ocean, and makes up a major portion of organic carbon in phytoplankton. Protein has long been thought to be highly labile in the environment and rapidly lost during diagenesis. However, the analysis of dissolved and particulate organic matter with NMR has revealed that much of dissolved and particulate marine organic nitrogen is linked by amide bonds, the very bonds that join amino acids in proteins. Throughout the global ocean, total hydrolysable amino acids (THAAs, the building blocks of proteins) can be measured in the water column and sediments, yet their biosynthetic source has remained elusive. Here, analytical techniques were developed combining protein solubilizing buffer extractions, gel electrophoresis, and proteomic mass spectrometry in order to investigate the biogeochemical significance of marine protein from primary production during transport and incorporation in sediments. These techniques enabled the detection and classification of previously unidentified marine sedimentary proteins. Specific proteins were tracked through the water column to continental shelf and deeper basin (3490 m)

sediments of the Bering Sea, one of the world's most productive ecosystems. Diatoms were observed to be the principal source of identifiable protein in sediments. *In situ* shipboard phytoplankton degradation experiments were conducted to follow protein degradation, and it was observed that individual proteins remained identifiable even after 53 days of microbial recycling. These studies show that proteins can be identified from complex environmental matrices, and the methods developed here can be applied to investigate and identify proteins in degraded organic matter from a broad range of sources. The longevity of some fraction of algal proteins indicates that carbon and nitrogen sources can be tracked down the marine water column to sediments in diatom-dominated systems as well as other types of phytoplankton. Using proteomic techniques to understand the marine carbon and nitrogen cycles will become increasingly important as climate change influences the timing, location, and phylogeny of those organisms responsible for oceanic primary production.

IDENTIFYING AND TRACKING MARINE PROTEIN AND ITS IMPORTANCE IN THE NITROGEN CYCLE USING PROTEOMICS

By

Elisha Kelly Moore

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Professor H. Rodger Harvey, Chair
Assistant Professor Nathan J. Edwards
Professor Lawrence W. Harding
Assistant Professor Johan Schijf
Associate Professor Marcelino T. Suzuki

Acknowledgements

I want to thank my advisor, Rodger Harvey, for his guidance, support and most of all his patience during my time as his graduate student. Rodger's cavalier spirit and endless enthusiasm for scientific exploration are a constant source of inspiration that I will treasure for the rest of my life. The advising, opportunities that Rodger gives his students to travel, and his international reputation have set a path for me towards an exciting and enlightening career. Rodger's curiosity and creativity in the world of marine organic geochemistry are second to none.

I thank my secondary adviser Brook Nunn for her proteomic expertise, skillful advising and significant contributions to this work. Brook is a true pioneer of marine proteomics. I greatly appreciate my committee members for their guidance and ability to use their broad skills and experience to improve my thesis. Contrary to popular belief, my project always made more sense after each committee meeting. Special thanks to the Goodlett Lab at the University of Washington for welcoming me into their lab group for each analytical visit.

I was blessed to work with a special group of people in Rodger's Marine Organic Geochemistry and Ecology Lab (MOGEL): Laura Belicka, Karen Taylor, Rachel Pleuthner, Jessica Faux, Angel Squire, Sejong Ju, and Abby Grabitz. From this group I would especially like to thank Laura, Karen, Rachel, and Jessica for their endless patience, troubleshooting help, research cruise guidance, humor, and grace. They have been wonderful role models throughout my graduate studies. Thanks to the crew and researchers of the R/V Knorr summer 2009 Bering Sea Ecosystem Study (BEST) cruise for a wonderful life changing experience.

I thank my many roommates for all of their patience and fun during my years of graduate school: Jason Edwards, Carlos Lozano, Karen Eisenriech, Lindsey Moore, Katherine Ziombra, Tammy Newcomer, Alison Zoll, Mike Seleckman, Emily Stefansen, Mindy Forsyth, Lisa Warden, Brian Laub, Owen McDonough, and Jake Hosen. Thanks to all of the students, faculty, and staff of Chesapeake Biological Laboratory and the Marine Estuarine and Environmental Science department of the University of Maryland for creating an exciting and collegial scientific community to study in.

I am eternally grateful to my loving family. The encouragement from across the country over the course of six years was a constant source of inspiration. I would not have made it through all of the analytical, physical, and emotional challenges if not for their guidance and love. I would like to thank my friends from around the country, especially those from Oregon: Kevin, Matt, Dylan, Seth, Carl, and Nick for all the fun and excitement during the many holiday and wedding visits to Oregon and New York. Special thanks to the Guilford Girls Terra Lederhouse, Lauren Culler, Nicole Harlen, and Kirsten Fristad for helping me transition to graduate school life.

Finally, I would like to thank my best friend and partner Lisa, her family, and her pets Mr. Bo Jangles, Isabel, and Mr. Rocky Balboa for making me happier than I ever thought possible. We zijn trekken hier weg om te winnen.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Due to its ubiquitous nature and complex chemistry, nitrogen has been described as an effective common constituent to study the components of the biosphere and the relations that exist between them (Boyd, 2001). One of the largest uncertainties in our understanding of the global nitrogen budget is the amount of reactive fixed nitrogen storage in most environmental reservoirs due to insufficient knowledge of ecosystem characteristics that control reactive nitrogen cycling and storage (Galloway et al., 2004 and references therein). Protein comprises the largest compartment of organic nitrogen in the ocean (Brown, 1991; Lourenco et al., 1998), and makes up a major portion of organic carbon in phytoplankton (Lee and Cronin, 1982; Wakeham et al., 1997). As much of the ocean operates under nitrogen limitation (Falkowski, 1997), fixed organic nitrogen is extremely important to microbial foodwebs. In addition, marine primary production remains a crucial global conduit for long-term atmospheric CO_2 drawdown and subsequent preservation (Nunn et al., 2010). Thus, as an important component of biomass and function in phytoplankton, protein can provide knowledge on the export and storage of both organic nitrogen and carbon.

Protein has long been considered to be highly labile in the environment and is rapidly lost during a complex series of reactions known as diagenesis (de Leeuw and Largeau, 1993). However, analysis of dissolved and particulate organic matter with NMR has revealed that much of the organic nitrogen is linked by amide bonds (McCarthy et al., 1997; Knicker, 2000; Zang et al., 2001), the very bonds that link amino acids together in proteins. The presence of these amide bonds suggest that the amino acid sequences of proteins may be intact. Throughout the global ocean total hydrolysable

amino acids (THAAs) have been measured in the water column and sediments (Hedges, 1991; Benner et al., 1992; Keil et al., 1994; McCarthy et al., 1998; Wakeham et al., 1997; Lee, 2000; Horiuchi et al., 2004) as a proxy for protein material. While amino acids are the building blocks of proteins, the biosynthetic source of sedimentary THAAs has remained elusive. The traditional acid hydrolysis technique of measuring THAAs in environmental samples for analysis of protein material destroys the amide bonds that link amino acids to build proteins. The inherent source and functional information encoded in the amino acid sequence is then lost. New approaches that take advantage of technological and computational innovations in proteomics from the biomedical field are needed to advance the study of protein and organic nitrogen in marine systems.

Before the late 1980s mass spectrometric analysis of large polar organic molecules was obstructed by the inability to vaporize such molecules into gas phase ions. Electrospray ionization overcame this obstacle by using high electric fields to desorb solute ions from small charged droplets of solution into an ambient bath gas, and has allowed for extensive analysis of proteins and other large polar molecules (Fenn, et al., 1989). The realization that automated comparison of MS measurements of peptides (protein fragments) to protein amino acid sequences available in databases could provide rapid and automated identification of proteins in laboratory samples, led to massive global expansion of the field of proteomics (Henzel et al., 1993; Eng et al., 1994; Shevchenko et al., 1996). Instrumentation and data analysis has advanced to the point that the mass spectra of a fragmented peptide can be interpreted to give the amino acid sequence of the protein. Using proteomic methods to answer fundamental geochemical questions has the potential to identify sedimentary marine protein and its fraction of

sedimentary organic nitrogen, identify protein sources from primary production and/or secondary consumers, and understand potential physical characteristics and/or modifications that allow protein or proteinaceous products to be preserved over long time scales.

Method synthesis

The potential wealth of source and functional information available in the form of intact proteins or peptides in marine sediments emphasizes the importance of method development in purifying these molecules for analysis. Proteomic mass spectrometry has allowed for rapid growth in the knowledge of protein structure and function in complex biological samples (Aebersold and Mann, 2003). However, in order to use these powerful techniques, sample proteins must be purified to the extent that unwanted sample interference materials do not obscure protein detection. The goals of this method survey were to execute and adapt existing methods to quantify, extract, purify, and characterize protein material from marine sediment. The methods that are developed will then be used to complement each other in the aim to track protein from the water column to sediments.

Methods of quantification that were tested include the Bradford assay, BCA assay, and total hydrolysable amino acids (THAAs). Various soft extraction buffers were tested and optimized for protein extraction from sediment samples. Precipitation, filtration, and ultrafiltration were used towards purification of protein extracts. Gel electro-elution and gel hydrolysis of 1D electrophoresis gels were used in concert with THAA analysis to characterize and quantify sediment protein extracts. Finally, gel

electrophoresis was used in an attempt to extract and purify proteins from cultured cells for analysis with proteomic mass spectrometry. The development of these methods laid the foundation for the studies in subsequent chapters.

Electrophoresis extraction and database evaluation

Extracting proteins from sediment has long been a challenge (Belluomini et al., 1986; Ogunseitan, 1993; Craig and Collins, 2000; Nunn and Keil, 2006). Acid hydrolysis is an effective method for extracting individual amino acids from sediment, however the source and functional information embedded in each protein's amino acid sequence is destroyed during the process. This method is not acceptable for analysis with proteomic mass spectrometry to identify the source and function of sample proteins. Potential interference materials and mechanisms present in sediment systems that can hinder extraction and identification of intact proteins and peptides include the binding of proteins to the mineral matrix (Mayer, 1994; Keil et al., 1994; Collins et al., 1995), organic matter co-extraction (Knicker and Hatcher, 1997), humic acids (Zang et al., 2000), algaenan (Nguyen and Harvey, 2003), and protein-protein aggregation (Nguyen and Harvey, 2001). To overcome these obstacles, various electrophoresis techniques were tested in order to solubilize proteins from marine sediment and then immobilize them in a polyacrylamide gel to wash away interference materials prior to digestion and analysis. The tests were made on Bering Sea surface sediment, and deep core sediment with the addition of bovine serum albumin (BSA) model protein. Multiple protein sequence databases were evaluated of varying complexity to see how the number of potential proteins in a database affected the number of proteins identified in an

environmental sample as described in chapter 3. The developed methods, database evaluations, and database selection will be used to characterize protein material from the marine water column to sediments.

Tracking diatom proteins from water column to sediments in the Bering Sea

With primary production rates up to 570 g C m^{-2} y^{-1}, the Bering Sea is perhaps the most productive region in the world (Sambrotto et al., 1986; McRoy et al., 1987; Walsh et al., 1989). Spring production is dominated by diatoms (Banahan and Goering, 1986; Springer et al., 1996), leading to carbon export flux up to 10 mmol C m^{-2} day^{-1} (Chen et al., 2003). This combination of high productivity of known algal communities and rapid transport to sediments make the Bering Sea an ideal system to study the early fate and export of algal proteins in marine systems. We have merged traditional protein buffer extraction and gel electrophoresis purification techniques with mass spectrometry-based proteomics in order to identify individual proteins from primary production (chlorophyll maximum particles), to sinking material (sediment traps), to surface sediments from the continental shelf and basin. By tracking proteins from algal bloom to sediments we have the opportunity to identify inputs to the sedimentary organic nitrogen and carbon pools and potential mechanism(s) which regulate the distributions observed during the initial stages of diagenesis.

Protein preservation over 11 and 53 day shipboard incubations

Diatoms are responsible for 25% of global primary production (Milligan and Morel, 2002). Diatom cell aggregates have been estimated to sink up to 100 m day^{-1} or

greater after bloom termination (Smetacek, 1985). In the highly productive (Sambrotto et al., 1986; McRoy et al., 1987; Walsh et al., 1989) diatom dominated Bering Sea (Banahan and Goering, 1986; Springer et al., 1996) it has been shown that algal proteins can be tracked from bloom to surface sediment (Moore et al., in review). This indicates that the cellular activities during bloom termination and the microbial recycling that follows are important to global organic nitrogen and carbon export. Diatom proteins observed over a 23 day lab based degradation experiment revealed trends in the rate of proteome restructuring and microbial recycling (Nunn et al., 2010). Recycling rates for identifiable proteins of Bering Sea bloom material were observed here to better understand these important environmental progressions.

Phytoplankton bloom material was collected in the Bering Sea during the spring of 2009 and 2010 and incubated under darkness in separate degradation experiments spanning 11 and 53 days respectively. The distribution of proteins was observed over the course of the incubations using shotgun proteomics, along with total hydrolysable amino acids (THAAs), total protein, particulate organic carbon (POC), particulate nitrogen (PN), and bacterial cell abundance. The observations seen here will give insight into the longevity of marine protein material and the mechanisms behinds its preservation.

Conclusions and implications

The work presented in this thesis represents an important step forward in the use of proteomics to characterize the marine organic nitrogen and carbon cycles. Source and functional information on algal material obtained will have increasingly important implications towards understanding the oceans biogeochemical cycles in changing global

climate. While much of observed primary production protein is recycled on short time scales, a fraction of this material remains identifiable for weeks to months and beyond. Marine proteomics will certainly be an important new addition to the scientific community from which a greater understanding of the ocean's biogeochemical cycles will be achieved.

Chapter 2: Evaluation of methods for sedimentary protein extraction and characterization

As the translated products of DNA, proteins are the functional bio-macromolecules in all organisms. The potential wealth of source and functional information available in the form of intact or partially degraded protein in marine sediment emphasizes the importance of sedimentary protein extraction method development to better understand the marine nitrogen and carbon cycles. Proteomic mass spectrometry has contributed to the rapid growth in the knowledge of protein structure and function (Aebersold and Mann, 2003). However, in order to use these powerful techniques, samples must be purified to the extent that sample interference materials do not obscure protein detection. Methods are needed to remove intact and/or partially degraded protein from complex sediment samples in a state that allows them to be observed by proteomic mass spectrometry. The goals of this method survey were to execute and adapt existing methods to quantify, extract, purify, and characterize protein material from marine sediment. The following methods were tested in order to accomplish these goals and arrive at the procedures that were used to analyze marine sedimentary and particulate proteins in subsequent chapters.

Total protein quantification

Total protein has been measured in a number of ways. Bradford, Lowry, and BCA assays are three common methods of measuring protein concentration in solution based on the absorbance of specific wavelengths of light (Bradford, 1976; Peterson, 1983; Smith et al., 1985). The Bradford and BCA assays were applied to sediment protein extracts to measure protein concentration prior to instrumental analysis.

Difficulties arose when trying to use these methods for measuring protein concentration in sediment extracts due to co-extracted sedimentary material and highly concentrated buffer components that increased the absorbance at each assay's specific wavelength. It was assumed that the same problem would arise for other methods that use light absorbance to measure protein concentration. These issues of poor accuracy and reproducibility made standard protein concentration assays unsuitable for protein quantification in sediment extracts.

Total hydrolysable amino acids (THAAs) are measured by hydrolyzing all protein material in a sample with a strong acid. The individual amino acids are then derivatized to allow detection with gas chromatography mass spectrometry (as described in detail on pages 60-61). The hydrolysis and derivatization process effectively removes sample interference materials giving accurate and quantitative measurements of each amino acid in a sample regardless of how complex the sample or extraction buffers may be. For these reasons THAAs were determined to be the suitable proxy for quantifying the concentration of total protein material. The EZFaast method by Phenomenex®, derivatization of AAs with propyl chloroformate and propanol for sensitive detection (Waldhier et al., 2010), was used because of its accuracy, speed and ease of use. Efforts were made to modify the EZFaast method to increase recovery of amino acids at low concentrations in protein extracts by increasing the volume of neutralized hydrolysate loaded onto EZFaast sorbent tips prior to derivatization. The loading volume was increased 2, 4, and 10 times greater than the standard volume (200 μl) from replicate sediment hydrolysates. Correct amino acid distribution was evaluated with hydrolysis and derivatization of bovine serum albumin (BSA). The mole % distribution of amino

acids was drastically changed as a result of the increased loading volume (Figure 2-1), possibly due to overloading of the sorbent tips. Correct amino acid distribution was accepted as more important than higher recoveries, and thus the EZFaast method was performed under the standard procedure.

Bulk extraction of sedimentary proteins

Extraction efficiency - Upon arriving at a suitable method to use as a proxy for total protein in complex sediment samples and extracts, optimization of protein extraction from sediment was undertaken to find the best extraction buffer and the best ratio of buffer to sediment sample for eventual proteomic analysis. Sodium hydroxide (NaOH) has been identified as an effective buffer for extracting THAAs from sediments with 60% recovery (Nunn and Keil, 2006). However, hydrolysis of extracted material by NaOH (Wu and Tanoue, 2002) makes it an unsuitable extraction buffer for proteomic analysis. To compare the THAA extraction efficiency of two softer extraction solvents, Bering Sea surface sediment was collected and separated into separate portions for THAA analysis and extraction.

One gram of whole un-extracted sediment was hydrolyzed in triplicate with 2 ml HCl for 4 hours (Cowie and Hedges, 1992). Hydrolyzed THAAs were then derivatized using the EZFaast method and quantified via GC/MS to measure total protein material and free amino acids in the sediment. Separate sediment portions were extracted with either 100 mM Ammonium bicarbonate or KS buffer (this buffer was developed as a protein extraction buffer by Kan et al. (2005) which we have termed the “kitchen sink” or KS buffer due to high concentrations of a suite of solubilizing agents including Urea,

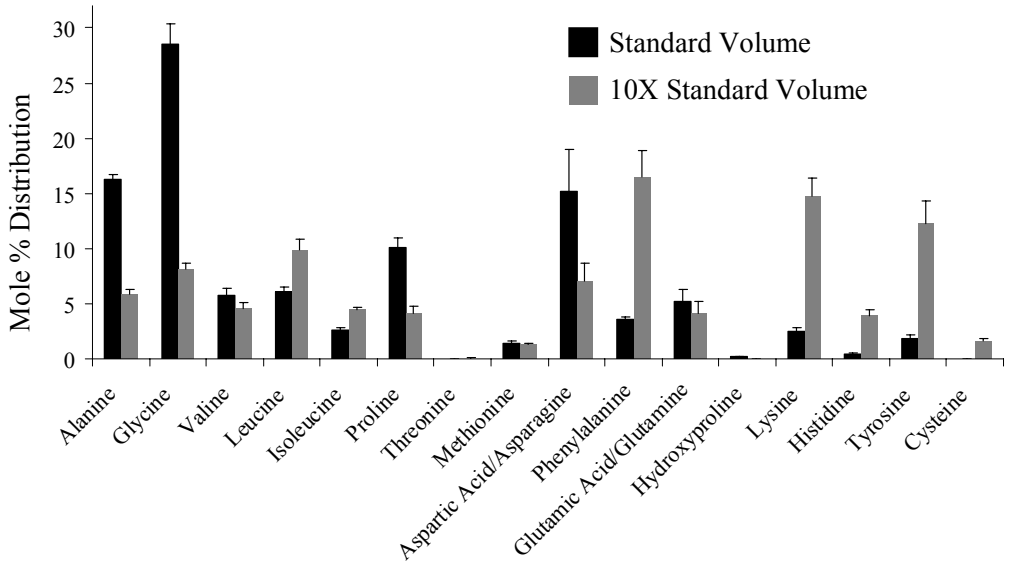

Figure 2-1. Amino acid distribution of sediment using the standard EZFaast sorbent tip volume (200 µl) and an adjusted volume 10 times greater than the standard volume. Amino acids are arranged in order of GC retention time. Error bars represent standard deviation.

Thiourea, detergents, etc.). Extractions were performed by sonicating 1 g of sediment in 2 ml buffer, centrifuging the sediment and buffer to remove particles, and decanting buffer extract from sediment. One ml of extracts and 1 g of extracted sediment were then hydrolyzed in triplicate with 2 ml HCl, derivatized using the EZFaast method, and finally analyzed via GC/MS to measure THAA content and calculate extraction efficiency of the KS buffer.

The THAA extraction efficiency was 11% for the KS buffer and 5% for Ammonium bicarbonate, low compared to ~60% THAA extraction efficiency observed using 0.5 M NaOH (Nunn and Keil, 2006). Comparison of amino acid distribution between sediments and extracts was used as a metric to evaluate whether each buffer extracts material that is representative of the extracted sediments. The individual amino acid distribution of extracted sediment closely resembled the amino acid distribution of un-extracted sediment for both extraction buffers, indicating that the amount of THAAs extracted by both buffers is relatively small compared to the THAAs of the whole sediment (Figure 2-2A). The amino acid distribution of the KS extract more closely resembles the amino acid distribution of the whole un-extracted sediment than does the amino acid distribution of the Ammonium bicarbonate extract (Figure 2-2B). This shows that the Kitchen Sink buffer obtains a more representative THAA distribution extract than Ammonium bicarbonate. Since the two extraction buffers obtain small fractions of the THAAs from the sediment this may indicate that different portions of protein materials are being extracted by the two buffers.

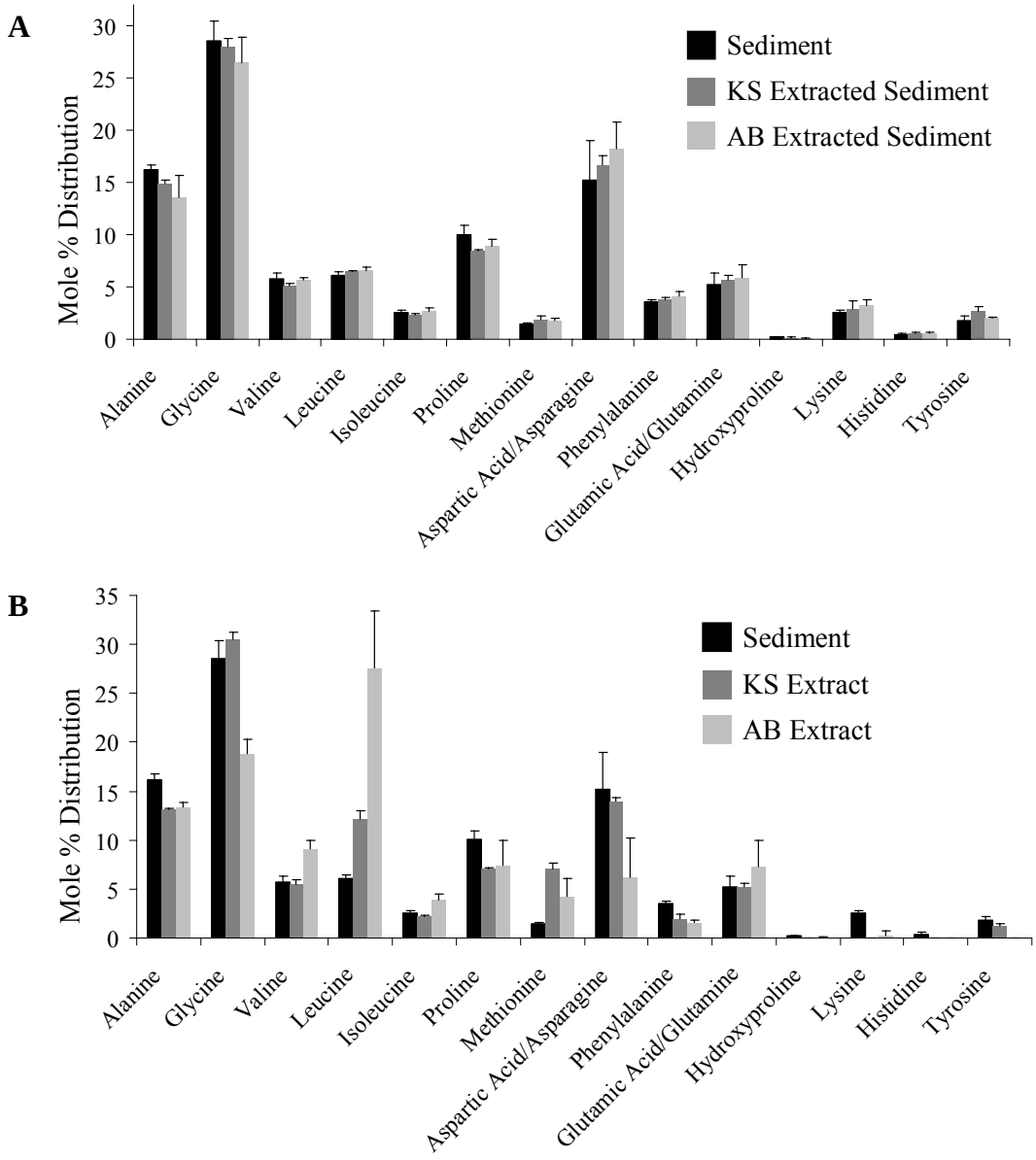

Figure 2-2. Amino acid distributions of (A) whole sediment vs. KS buffer extracted sediment and Ammonium bicarbonate extracted sediment (extracted sediment's THAA distributions are measured on sediment after the buffer extract is removed); (B) whole unextracted sediment vs. KS buffer sediment extract and Ammonium bicarbonate sediment extract (sediment extract's THAA distribution are measured on the buffer extract that is removed from sediment). Error bars represent standard deviation.

Extraction optimization – After initial buffer extraction comparisons, separate sediment portions were then extracted by Kitchen Sink and Ammonium bicarbonate buffers in five different buffer to sediment volume:volume ratios (1:1, 2:1, 3:1, 5:1, and 10:1) to identify the most efficient ratio. Ten separate 2 ml volumes of sediment were placed in falcon tubes. In each tube 1 mg of Cytochrome C and 1 mg of bovine serum albumin (BSA) was added to increase the amount of potentially extractable protein. The sediments and protein additions were vortexed to mix uniformly. Separate buffer volumes were then added to each tube (Table 2-1). Tubes were vortexed again to mix buffer with sediment and samples were pulse sonicated for 1 minute on ice. The tubes were then centrifuged for 1 min at 1000 rpm and the extracts were pipetted from the sediments. Extracts were hydrolysed and then analyzed for THAAs via GC/MS. The highest total amount of THAAs extracted per gram of sediment was the 10:1 ratio for both buffers (Table 2-1). The 5:1 buffer ratio was chosen for future procedures as it was nearly equal to the 10:1 in yields of THAA's extracted per gram of sediment and allowed lower THAA concentration to be followed in dilute environmental samples. Thus, the 5:1 ratio yields the best combination of high THAA content and small buffer volume. Extraction recoveries were higher for the ratio optimization than previous experiments due to the Cytochrome C and BSA added to sediment. This additional protein also likely resulted in a smaller difference in extraction efficiency between KS buffer and Ammonium bicarbonate, although the KS buffer was still higher.

Table 2-1. Results of buffer:sediment (*volume:volume*) extraction experiment. Approximately 2 ml of sediment was used in each extraction. Mass units for sediment are all in wet weight. % Recovery = total amount of THAAs extracted from THAAs in sediment.

Extraction Buffer	Mass Sed (g)	Buffer volume (ml)	ug THAA extract/ ml buffer	ug THAA extracted/ g sediment	% Recovery
Kitchen	2.77	2	327.5	236.4	11.0
Sink	2.97	4	237.3	319.5	15.1
	2.94	6	173.6	354.2	16.7
	2.92	10	143.6	491.9	23.1
	2.94	20	74.1	504.0	23.7
Ammonium	2.76	2	356.0	258.0	12.0
Bicarbonate	2.87	4	204.7	285.3	13.4
	2.86	6	154.6	324.3	15.2
	2.99	10	137.6	460.1	21.7
	3.04	20	70.1	461.2	21.8

Protein extract purification

The issue of diluting extracted protein in large extraction buffer volumes prompted the search for a method to concentrate and purify protein from the extraction buffer. Trichloroacetic acid (TCA) and chloroform methanol water (CMW) precipitations were tested on separate 100 μg/ml Cytochrome C and BSA solutions in nanopure water and 3 micron filtered seawater. The maximum amount of protein able to be recovered from nanopure water was only 33%, and filtered sea water only 11% for either precipitation method. The low recoveries and concerns with resolubilization of protein precipitates for proteomic analysis led to other concentration methods to be explored.

Extract filtration – Nylon GD/X syringe filters, PVDF Sterivex cartridge filters, and PES Sterivex cartridge filters (all 0.2 μm) were all tested to accumulate protein to potential filter binding sites and remove co-extracted material, followed by elution with Acetonitrile. BSA was diluted in nanopure water to a concentration of 431 μg protein/ml and initially passed through the GD/X filter. The protein concentration of the filtrate was measured with fluorescent absorbance to observe how much protein was bound to the filter. Approximately 50% of protein was bound to the filter from the first ml of solution filtered (Figure 2-3). The amount of protein bound from subsequent filtered solution declined as more solution passed through the filter indicating the binding sites were being occupied. Protein solutions of varying concentrations were also passed through each of the Sterivex cartidge filters and very minimal protein binding was observed. A 20 mg/ml protein solution was needed for the Sterivex filters to demonstrate any protein concentration above 50% from the solution. As 20 mg/ml is a very unrealistic protein

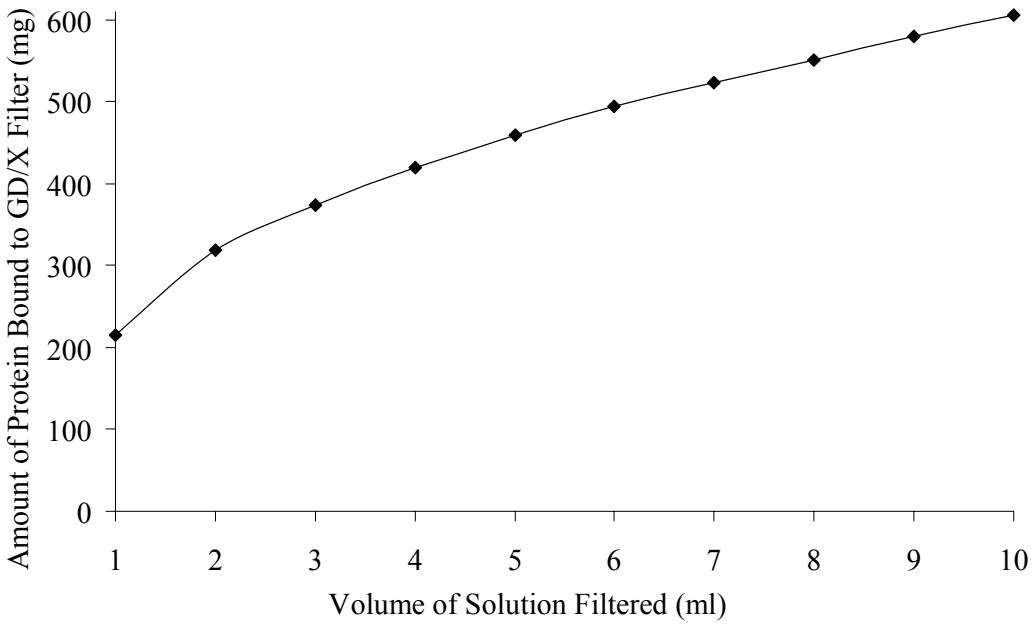

Figure 2-3. Amount of protein (BSA) bound to GD/X nylon filter from increasing volume of protein solution passed through the filter.

extract concentration, the Sterivex filters were abandoned and only the GD/X filter was tested for protein binding from sediment extracts.

Approximately 2 ml of Bering Sea surface sediment was extracted with 10 ml of KS buffer. Four ml of extract was passed through a GD/X filter. The filter was then eluted three times with the same 1 ml volume of acetonitrile which was then hydrolyzed, derivatized and analyzed for THAAs. The amount of THAAs recovered from the GD/X filter was compared with the THAA concentration of the total extract, and only 0.8% of THAAs from the original extract was recovered. The combination of the protein solubilizing KS buffer and coextracted sediment material filling nylon binding sites likely limited the amount of protein material collected by the filter. For these reasons, filters were abandoned as a method for concentrating protein from sediment extracts.

While GD/X and Sterivex cartridge filters appeared to be ineffective at quantitative protein concentration, centrifuge filters remained a possibility for qualitative analysis of extracted protein material. Ultrafilters have been used to crudely separate protein molecular weight fractions from sediments (Pantoja and Lee, 1999; Nunn and Keil, 2005) followed by THAA analysis. We were interested if this technique could be used to separate protein material prior to LC/MS analysis. Cytochrome C, used as a model protein, was added and vortexed with estuarine sediment and ultrafiltered to attempt to purify protein from sediment extracts. Protein material was extracted from the sediment using either KS buffer or 6 M urea. Extracts were separated using 3000 Da MW microcon ultrafilters and rinsed five times with 44 mM Ammonium bicarbonate to desalt and clean up the extracts. The >3000 Da MW extracts were digested with trypsin overnight and analyzed using LC/MS. No Cytochrome C peptides were identified

reproducibly from any of the extracts. Binding of Cytochrome C to the filter and matrix material likely blocked identification by LC/MS analysis and proteomic database searching. Furthermore, in an environmental sample partially degraded proteins with a molecular weight <3000 Da would be lost through the filter. While ultrafilters may be useful for total hydrolysable amino acid analyses, their potentially high binding of extracted proteins and potential loss of low molecular weight material found them to be unsuitable for purposes of qualitative proteomic characterization. This led to the testing of other purification methods.

Purification and characterization from bulk extracts

Gel electrophoresis molecular weight separation – After the unsuccessful use of ultrafilters to identify model proteins with proteomic mass spectrometry, gel electrophoresis was used as a molecular weight characterization technique prior to quantitative and proteomic analysis. Sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE) separates proteins based primarily on their molecular weights (Laemmli, 1970). SDS binds to hydrophobic portions of a protein, disrupting its folded structure and allowing it to exist stably in solution in an extended conformation. As a result, the length of the SDS-protein complex is proportional to its molecular weight. The ease of execution and wide application of SDS-PAGE have made it in an important analytical technique in biochemical research.

To attempt to measure the amount of THAAs in different gel molecular weight sections, gel hydrolysis and electro-elution were tested. For gel hydrolysis, KS buffer extract of Bering Sea surface sediment was loaded onto three 1-D 12% Bis-Tris gels and

run for 1 hour at 180 volts (dark streaks of sediment material were observed down the length of the gel). Each lane was cut into five molecular weight sections: <10 kDa; 10-25 kDa; 25-50 kDa; 50-100 kDa; >100 kDa; top portion of the gel containing material that "did not transit" (DNT). Gel sections were then submerged in 6 N HCl for 4 hours at 105^{o}C. The hydrolysates were then analyzed for THAAs using the EZFaast method and GC/MS. Amino acid signals from the gel hydrolysates were indistinguishable from blanks, and the gel hydrolysis method was deemed ineffective.

Electro-elution is a method that elutes protein material from gels the same way it is mobilized into the gel, by using an electric current. The eluted material is then collected in buffer for subsequent analysis. Electro-elution can be used on small excised sections from an electrophoresis gel, allowing analysis of material in a specific molecular weight range. To test electro-elution to isolate molecular weight distributions of protein extracts from a biological sample, separate portions of diatom culture, *Thalassiosira weisflogii*, were extracted using ammonium bicarbonate and KS buffer. Sections from the diatom extract gel lanes were excised and electro-eluted (protein material mobilized out of gel sections using electrophoresis voltage) for THAA analysis to measure total protein material from different molecular weight sections. The excised molecular weight ranges were the same as the gel hydrolysis sections above (Figure 2-4). An advantage of the ammonium bicarbonate extracts is that they can be concentrated to a greater extent via speedvac than KS extracts. This allows for greater THAA loading per gel lane with ammonium bicarbonate extracts.

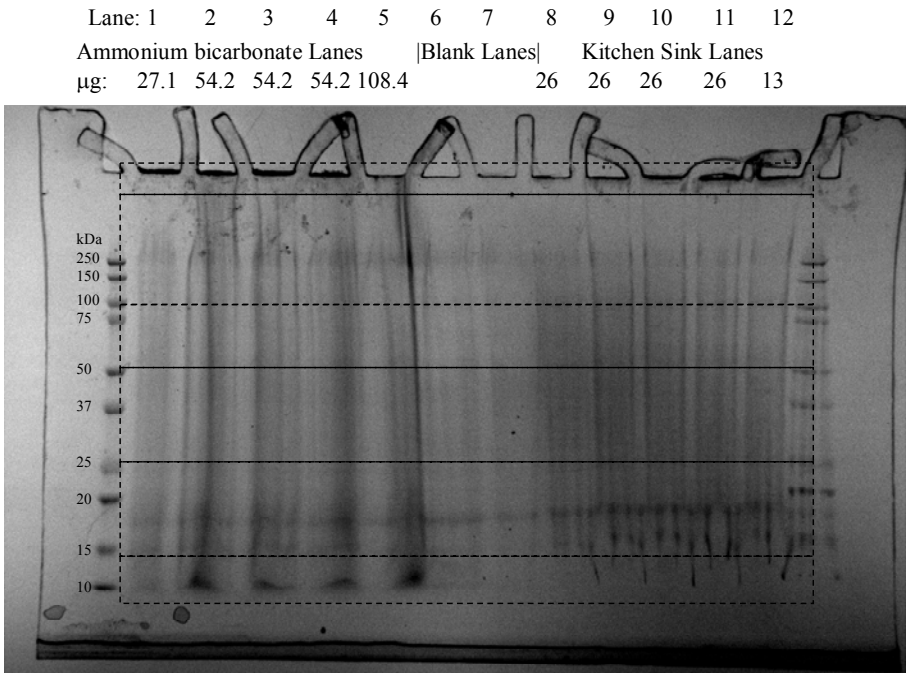

Figure 2-4. 1-D electrophorsis gel of Ammonium bicarbonate and Kitchen Sink extracts. Numbers above gel lanes indicate the amount of THAAs loaded into each lane. Lines across gel indicate molecular weight ranges where gel sections were excised for electro-elution.

The gel sections from lanes 2, 3, and 4 were excised and electro-eluted using 1 M Ammonium bicarbonate and 1% SDS buffer for 5 hours. Electro-elution of lanes 3 and 4 were concentrated 2 fold under nitrogen to increase the overall THAA concentration. Individual amino acid amounts per gel section for lane 2 (unconcentrated) ranged from 0.034 μg for Alanine in the 50-100 kDa section up to 2.188 μg for Aspartic Acid in the DNT section. The sum of the THAAs eluted from all the gel sections of lane 2 (unconcentrated) was 19.7 μg or 36.2% of the 54.3 μg of THAAs loaded onto the gel. The THAAs eluted from the gel sections of lane 3 and 4 (concentrated) were 11.2 μg or 20.5% recovery and 6.9 μg or 12.7% recovery. Blanks used from gel lanes that were not loaded with protein in a separate gel that was run simultaneously and accounted for 14% of measured THAAs (reported values from sample lanes are blank corrected). Low recoveries from sample lanes may have led to high THAA proportions in blank lanes. Unexpectedly, the two lanes of concentrated electro-elutions (lanes 3 and 4) had much lower amino acid amounts than the un-concentrated electro-elutions. This may have been caused by extract condensation to the inside of the concentration tube.

The percent contribution of the different molecular weight ranges to total electro-eluted THAAs varied somewhat between the three diatom extract gel lanes (Figure 2-5). The largest difference observed was the DNT molecular weight range, in which lanes 2 and 4 were similar, but lane 3 was very high. Despite the differences the DNT and 25-50 kDa molecular weight sections were the highest contributions for all other molecular weight ranges of the three lanes followed by 10-25 kDa, >100 kDa, <10 kDa, and finally 50-100 kDa. These results showed that electro-elution could be an intriguing method for extracting THAAs from different molecular weight ranges of a 1-D gel, however small

protein gel loading limits and low signal to blank ratios represent problems to reproducibility. Only 5 amino acids could be consistently measured from gel electro-eluates, resulting in highly variable amino acid distributions.

The greatest unresolvable issue with gel electro-elution was glycine contamination. Glycine is a major component of gel running buffers, and despite using different combinations of Bis-Tris gels, Tris-HCl gels, XT-MOPS running buffer, and Tris-Tricine running buffer, the glycine signal could not be avoided. While the electro-elution method was eventually abandoned, the gel/buffer combination tests did reveal that Bis-Tris gels with XT-MOPS running buffer was the best combination for electrophoretic purification of proteins in extracted sediment material.

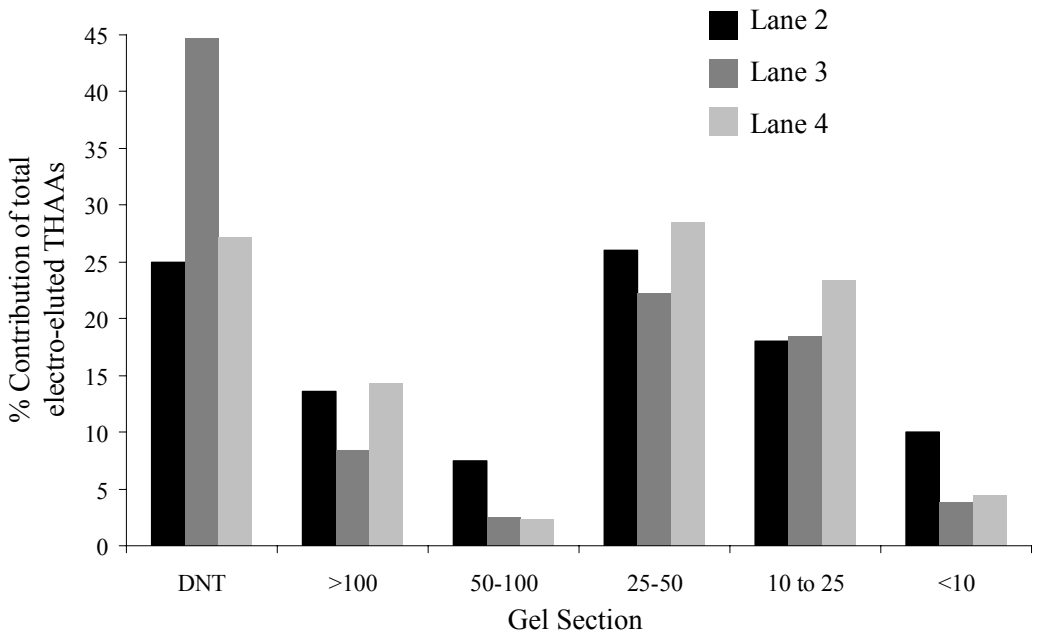

Figure 2-5. Percent contributions of THAAs across different molecular weight ranges obtained by electro-elution of gel sections for lanes 2, 3, and 4 of the diatom extract 1-D gel shown in Figure 2-4.

Combined protein extraction purification and characterization in Bering Sea suspended particles and sediments

Electrophoresis extraction - During the many gel separations of sediment extracts a common trend became apparent. There was more visible/stainable protein-like material in a gel lane loaded with more particulate material compared to a gel lane that was loaded with less particulate material from the same sediment. In Bering Sea continental shelf sediment collected in 2008 this included both a visible band at approximately 50 kDa and low molecular weight material at the bottom of the gel lanes (Figure 2-6A). In the gel lanes that were loaded with extract from which particulate matter was excluded, there are no visible protein bands and very little low molecular weight material at the bottom of the gel lanes (Figure 2-6B). This would suggest that some intact protein is associated with particulate material, and that proteins were extracted from that particulate material by the electrophoretic process.

To further investigate the prospect of electrophoresis extraction of proteins from biological samples, the diatom *Thalassiosira weissfloggii* was cultured and intact cells and cell extracts were loaded onto a 1D gel with an extraction step and without a prior extraction step. The goal was to compare protein extraction between a more conventional method of extraction using solvents with a novel method of electrophoresis extraction.

Diatom cells were concentrated from culture by centrifugation at 10,000 rpm for 5 minutes and separated into two different Eppendorf tubes. In one tube 25 µl of gel XT-MOPS running buffer was added to 92.5 mg of diatom cells and sonicated (30 seconds, 10% duty cycle, output control of 2). After sonication the cell/buffer slurry was loaded directly onto a 12% Bis-Tris gel with 0.02 g loaded into lane 2 and 0.01 g loaded into

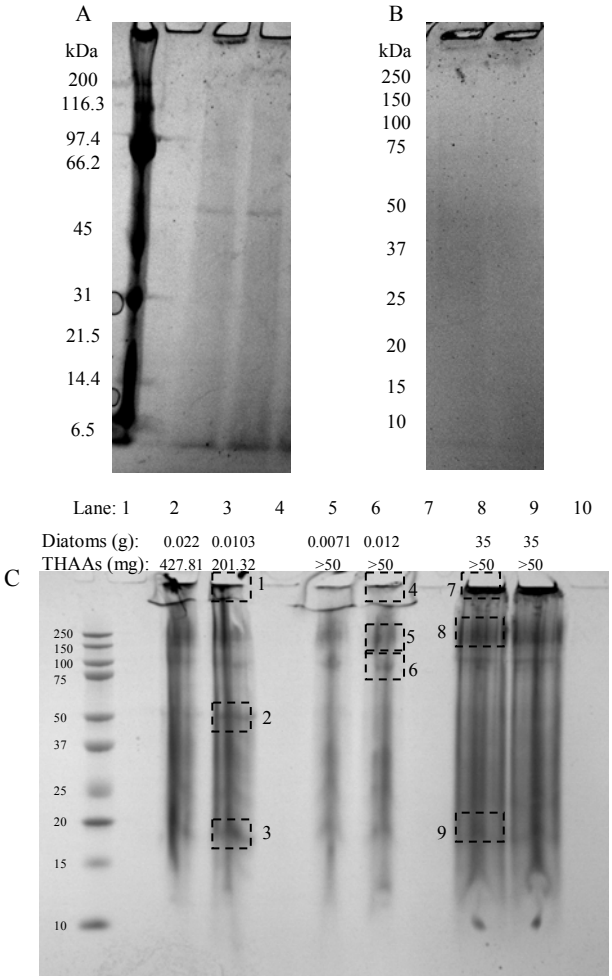

Figure 2-6. (A) Comparison of 1D 12% Bis-Tris gel separation of Ammonium bicarbonate extract from Bering Sea sediment with some particulate matter included and 10 μg THAA per lane vs. (B) Bis-Tris gel of Ammonium bicarbonate extract from Bering Sea sediment with particulate material excluded and >10 μg THAA per lane; (C) Image of 12% Bis Tris gel loaded with sonicated diatom cells in lanes 2 and 3, extracted diatom cells in lanes 5 and 6, and diatom cell extract in lanes 8 and 9. Mass of diatom cell debris and THAAs listed above each lane, dashed rectangles indicate excised gel sections for analysis, sections 8 and 9 indicate molecular weight sections digested and analyzed from precious KS diatom extract gels.

lane 3, which corresponds to 427.81 μg and 201.32 μg of THAAs respectively (Figure 2-6C). The sonicated cells were loaded with a spatula, which was rinsed with gel buffer to remove cellular material and wash it into the gel lane. In the second eppendorf tube 100 μl of kitchen sink buffer was added to 94.5 mg of diatom cells and sonicated (same procedure as above) to extract protein material. The extracted cell debris was centrifuged to the bottom of the eppendorf tube and the overlying extract pipetted from the cells. Thirty five μl of supernatant were loaded into two gel lanes each (lanes 8 and 9), and 7.1 mg and 12 mg of cell debris were loaded into two separate gel lanes (lanes 5 and 6) with a metal spatula (Figure 2-6C).

The 12% Bis Tris SDS PAGE was run for 68 minutes at 180 volts until the ion front (bottom edge of material moving through gel) traveled to the bottom of the gel. Gel sections were excised and destained twice for two separate 30 minute intervals using Sigma Co. destaining solution at 37^{o}C. After destaining, the gel sections were dried under nitrogen for 15 minutes and digested separately in eppendorf tubes with 1 μg of Trypsin in a 40 mM NH_4HCO_3/9% Acetonitrile solution at 37^{o}C for 5 hours. Gel bands at approximately 200 kDa and 20 kDa in the diatom extract gel lanes were not excised because they were already known to contain RuBisCO and chlorophyll binding proteins respectively from previous gel digests.

The gel digests were analyzed on an Agilen LC/MS (electrospray ion trap) to identify potential peptide masses and fragmentation masses. The peptide masses and fragmentation mass spectra were then searched against the MSDB database using Mascot (Perkins et al., 1999) to identify potential protein sequence matches with >90% confidence (Table 2-2). The DNT (do not transit) sections of lane 8 and lane 3 both

Table 2-2. Peptides and proteins identified from digests of electrophoresis extraction experiment on diatom cells. The section number in parenthesis is corresponds to gel section numbers in Figure 2-6C. ID'ed mass is the mass (Da) of the identified protein.

	Lane 3	Lane 6	Lane 8
Section (#)	DNT (1)	DNT (4)	DNT (7)
Peptide(s)	TFQGIATGIIVER	ADTRNAIELLR	NGALDFGWDSFDEETK
Protein	RuBisCO, large subunit	Putative DNA mismatch repair protein	Fucoxanthin chlorophyll a/c-binding protein precursor
Organism	Asparagopsis armata	Staphylococcus saprophyticus	Cylindrotheca fusiformis
ID'ed Mass (Da)	25,024	63,296	21,217
Section	50 kDa (2)	150 kDa (5)	DNT (7)
Peptide(s)	DTDVLALFR VALESMILAR	DTDVLALFR TFQGIATGIIVER	DTDVLALFR GGLDFLKDDENINSQPFMR
Protein	RuBisCO, large subunit	RuBisCO, large subunit	RuBisCO, large subunit
Organism	Nemalionopsis shawii	Mazzaella japonica	Hypnea cornuta
ID'ed Mass (Da)	50,176	25,024	50,176
Section	18 kDa (3)	150 kDa (5)	150 (8)
Peptide(s)	NGYIDFGWDDFDEETK IAQLAFLGKIVTR	NGYIDFGWDDFDEETK	TFQGNATGIIVER
Protein	Fucoxanthin-chlorophyll a/c light-harvesting protein	Chloroplast fucoxanthin chlorophyll a/c-binding protein	RuBisCO, large subunit
Organism	Skeletonema costatum	Cylindrotheca fusiformis	Hypnea cornuta
ID'ed Mass (Da)	15,977	22,664	50,176
Section	18 kDa (3)	75 kDa (6)	18 kDa (9)
Peptide(s)	NGALDFGWDSFDEETK	no IDs	NGYIDFGWDDFDEETK
Protein	Fucoxanthin chlorophyll a/c-binding protein		Chloroplast fucoxanthin chlorophyll a/c-binding protein
Organism	Cylindrotheca fusiformis		Cylindrotheca fusiformis
ID'ed Mass (Da)	15,672		22,664

contained RuBisCO, which was not identified in the DNT section of lane 6. The diatom cell lanes and the extracted diatom cell lanes both contained RuBisCO and fucoxantin chlorophyll binding protein. Results show that protein can be extracted directly from sonicated diatom cells by electrophoresis without an extraction step to remove proteins from cellular material.

To test the various methods described in this chapter, an extraction experiment was conducted comparing buffer extractions with KS buffer, Ammonium bicarbonate, NaOH, electrophoresis extraction, and protein precipitation. The methods were all tested on estuarine sediments with added diatom culture cells. Gel sections or precipitations were each digested with trypsin and analyzed with an Agilent LC/MS ion trap. No diatom peptides or other environmental peptides were identified from any of the extraction methods. This experiment in which protein extraction and identification was unsuccessful led to the utilization of more refined proteomic techniques including improved protein digestion with reduction and alkylation steps (Schevchenko, 1996), and more sensitive instrumentation. These refined techniques were applied to Bering Sea sediments.

Method testing on Bering Sea suspended particles and sediment

To test the electrophoresis extraction/purification methods along with THAA quantification on environmental samples and track the fate of marine protein from the water column to the sediment, suspended particulate material from the water column and surface sediment were collected during the Bering Ecosystem Study (BEST) spring and summer 2008 cruises. Surface water particles (POM) was collected in spring 2008, POM

from waters below the chlorophyll maximum was collected in summer 2008, and surface sediment was also collected in summer 2008 at the same time as POM below the chlorophyll max. All samples were collected from the same location (57.9003N, 169.2318W). Particles were collected with combusted 47 mm glass fiber filters, and sediment cores with a multicorer.

Sediment and POM samples were extracted using KS extraction buffer. POM extracts and sediment/extraction buffer slurry were loaded directly onto 1D electrophoresis gels for protein purification. The electrophoresis gels were run for a short amount of time (15 min) to mobilize protein into the gel but not undergo molecular weight separation. Protein sections were excised from the gel and digested with trypsin. Replicate gel sections for each sample were excised, electro-eluted, and analyzed for THAAs. The THAA amounts for each gel sample were used to adjust the concentration of the gel digests via dilution as necessary so that the THAA concentrations would be equal for all the samples prior to LC/MS analysis (for description of LC/MS parameters see Nunn et al., 2010). The resulting mass spectra were searched with SEQUEST against a database containing the proteomes of *Thalassiosira pseudonana* (marine diatom), *Prochlorococcus marinus* (marine cyanobacteria), and *Pelagibacter ubique* (SAR11 bacteria). Probabilistic scoring was given by PeptideProphet and ProteinProphet. Peptide and protein identifications with statistical confidence >90% were accepted as positive detections.

The majority of the peptides identified in each of the samples were correlated to *T. pseudonana* protein amino acid sequences. The number of peptides correlated to *T. pseudonana* proteins increased down the water column to the sediment (Table 2-3).

Table 2-3. Number of proteins identified by organism from Bering Sea suspended particles and surface sediment: *T. pseudonana*; *P. ubique*; *P. marinus*.

Sample	Collection Date	*T. pseudonana*	*P. ubique*	*P. marinus*
Surface	4/18/2008	7	4	0
Below Chl-Max	7/14/2008	16	3	1
Sediment	7/14/2008	34	0	0

The number of peptide sequences correlated to *P. marinus* and *P. ubique* proteins identified in the surface was almost the same as below the chlorophyll max, but there were no peptides correlated to *P. marinus* or *P. ubique* proteins identified in the sediment. One *P. marinus* protein was identified below the chlorophyll maximum. *T. pseudonana* proteins located in the chloroplast, mitochondria, secretory, and uncharacterized compartments of the cell were all identified in the sediment. Different proportions of proteins, calculated from the number of proteins identified from a certain cellular location, were observed as the diatoms moved down the water column to the sediment (Figure 2-7). For diatom proteins, the proportion of chloroplast proteins relative to other proteins increased down the water column to the sediment. This leads to the hypothesis that chloroplast proteins may be preferentially preserved during the short term degradation of diatom cells.

A larger number of proteins were identified in the sediment than in the surface water or below the chlorophyll max (Table 2-3). This is unexpected and possibly due to the timing of sample collection, which did not occur during the peak phytoplankton bloom period, and the higher concentration of diatom cells to other cells in the sediment because of differential sinking rates. The surface water POM samples were likely collected before the spring bloom was at its highest state, while the POM sample below the chlorophyll maximum may have been collected after much of the spring bloom material below the chlorophyll max had already sank to the sediment. Because diatoms are relatively heavy (Dunne et al., 2005) they can sink during their life cycle, and accumulate in the sediment to a much greater extent than other organisms that originate in the water column, enhancing their protein contribution to the sediment.

Figure 2-7: Cellular compartmentalization of diatom proteins detected in surface water POM, water column POM below the Chl-Max, and surface sediment samples.

Further analysis of water column particles would be needed to ensure that phytoplankton proteins were not being missed during analysis.

Results suggest that protein remains identifiable during and after transport through the water column to the sediment. Second, the dominant identifiable source of protein to the sediment was primary production over bacteria, suggesting that larger databases containing more bacterial proteomes should be tested. Third, the proportion of diatom chloroplast proteins increased down the water column which could suggest that there is selective degradation of proteins. There were relatively low numbers of proteins identified in the surface water and in the water column below the chlorophyll maximum when one considers the diversity of the marine microbial community and abundance of phytoplankton cells in the Bering Sea. This is likely due to inefficient collection of bacterioplankton, limited size of the protein sequence database used, and loss during electrophoresis and subsequent sample processing. Further work would be needed to better track the sources and inputs of protein from the water column to the sediment using larger protein databases containing many proteomes that reflect the high microbial diversity of the Bering Sea.

Chapter 3: Evaluation of electrophoretic protein extraction and database-driven protein identification from marine sediments

Abstract

Proteins comprise a major component of organic carbon and nitrogen produced globally, and are likely an important fraction of organic matter in sediments and soils. Extracting the protein component from sediments and soils for structural characterization and identification represents a substantial challenge given the range of products and functionalities present in the complex matrix. Multiple forms of gel electrophoresis were evaluated as a means of enhancing recovery of sedimentary protein prior to proteomic characterization and compared with a direct enzymatic digestion of proteins in sediments. Resulting tryptic peptides were analyzed using shotgun proteomic tandem mass spectrometry and evaluated with SEQUEST, PeptideProphet, and ProteinProphet. Multiple databases were tested to examine the ability to confidently identify proteins from environmental samples. Following evaluation of electrophoretic extraction of proteins from sediments more tests were completed to optimize the recovery of an experimentally added standard protein (BSA) in older (>1ky) sediments. Extractions of the protein component from sediments via direct electrophoresis of a slurry mixture of sediments and the specified extraction buffer resulted in the greatest number of confident protein identifications and highest sequence coverage of the BSA standard. Searching tandem mass spectra against larger databases with higher diversity of proteomes did not yield a greater number of, or more confidence in, protein identifications. Regardless of the protein database used, identified peptides correlated to proteins with the same function across taxa. This suggests that while determining taxonomic-level information remains a challenge, it is possible to confidently assign the function of identified proteins.

Introduction

Proteins make up the majority of organic nitrogen in marine phytoplankton (Lourenco et al. 1998) and total hydrolysable amino acids (THAAs) account for up to 30-40% of particulate nitrogen in marine sediments (Cowie and Hedges, 1992a; Grutters et al., 2001). Acid hydrolysis has been used for decades to assist in the extraction of amino acids from sediment (Lee and Cronin, 1982; Hedges, 1991; Benner et al., 1992; Keil et al., 1994; McCarthy et al., 1998; Horiuchi et al., 2004), but this destroys the peptide bond and consequently the primary sequence of proteins. While total amino acids can provide a proxy of total protein material, the functional and source information embedded in each protein's amino acid sequence is lost. To fully characterize the cellular machinery of organisms responsible for the biogeochemical cycles of nitrogen and carbon, the identification of peptides and/or proteins in marine sediment is needed.

Extracting proteins from sediment has long been a challenge (Belluomini et al., 1986; Ogunseitan, 1993; Craig and Collins, 2000; Nunn and Keil, 2006). The potential interferences present in sediment systems include the binding of proteins to the mineral matrix (Mayer, 1994; Keil et al., 1994; Collins et al., 1995), organic matter co-extraction (Knicker and Hatcher, 1997), humic acids (Zang et al., 2000), algaenan (Nguyen and Harvey, 2003), and protein-protein aggregation, which limit solubility (Nguyen and Harvey, 2001). These varied interactions have lead to the need for strong solubilizing agents when attempting to retrieve and isolate the protein component. Although the application of strong agents to solubilize proteins can be effective, it results in the co-extraction of a suite of unknown compounds with similar physiochemical properties as protein from the sediment or soil matrix. These mixtures are inherently complex and

interfere with the purification and identification of peptides and proteins (Cheng et al., 1975; Limmer and Wilson, 1980; Nunn and Timperman, 2007).

Gel electrophoresis has been widely used for decades as a protein separation and visualization technique. Sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE) and related approaches separate proteins based primarily on their molecular weights (Laemmli, 1970). The wide application of SDS-PAGE and its ability to solubilize and immobilize proteins have made it a standard analytical technique for protein separation and isolation across the fields of biochemistry, cell biology, and medical sciences (Reisfeld et al., 1962; Laver, 1964; Shapiro et al., 1967; Fairbanks et al., 1971; Maizel, 2000; Pederson, 2008). The focus here was to develop and validate a modified electrophoretic approach as an extraction and preparative technique prior to high performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS) analysis. This methodology is founded on the excision of protein bands from electrophoresis gels, followed by enzymatic digestion, a frequently used method for protein identification using HPLC-MS/MS (Hirano et al., 1992; Schevchenko et al., 1996; Kuster, et al., 1998).

Ongoing advancements in proteomic use of HPLC-MS/MS have increased sensitivity and detection limits, providing the user with an increased ability to identify peptides from complex mixtures (Schulze et al., 2005; Morris et al., 2010; Dong et al., 2010). A caveat, however, is that such techniques also require that protein samples are free from interfering substances, including salts, detergents, or humic acids. Such contaminants have both practical (clogged chromatography columns or electrospray needles), and forensic impacts (decreased signal to noise ratio, or competition with true

peptide signals) that compromise analysis. The goals of this work were thus twofold: the first goal was to optimize the extraction of proteins from marine sediments. For this purpose we evaluated two methods that involved an SDS-PAGE clean-up step: 1) a more traditional method where the buffer-solubilized material is separated from the particles and loaded directly onto gels, and 2) a slurry extraction method where the buffer-solubilized material remains with the sediment particles and is loaded together onto gels. Electrophoresis gels investigated included preparatory tube gels and standard 1-dimensional flat gels. In addition, multiple combinations of extraction buffers were tested. The second goal was to assess the effectiveness of proteomic database complexity for searching against mass spectra from environmental samples. Mass spectra were searched against five databases of varying size, in terms of the number of protein sequences, to evaluate database-driven protein identifications using probabilistic scoring. Continental shelf sediments from the Bering Sea were used as the test sample since this area is one of the world's most productive ecosystems (Sambrotto et al., 1986; McRoy, 1987; Walsh et al., 1989), and is known to be diatom dominated during spring blooms. High carbon export flux (Chen et al., 2003) in the spring leading to high sedimentary biomass (Grebmeier et al., 1988) makes it a useful system to explore sedimentary protein extraction and evaluate information from multiple database searches.

Materials and Procedures

Protein extraction using buffers – Bering Sea surface sediments were extracted using a buffer followed by SDS-PAGE. Extraction buffer was prepared prior to sediment additions: 7 M urea, 2 M thiourea, 0.01 M Tris-HCl, 1 mM EDTA, 10% v/v glycerol, 2% CHAPS, 0.2% w/v ampholytes, 2 mM Tributyl-phosphine (Kan et al., 2005). The mixture includes chaotropic agents, detergents, denaturants and salts, thus proteins are solubilized and stabilized while avoiding degradation. The use of strong chaotropic agents and subsequent trypsin digestion alleviated the need for protease inhibitors. Replicate aliquots of each treatment were used for amino acid analyses to measure recoveries.

For the traditional method, approximately 1.5 g dry weight surface sediment (~1 ml wet sediment) was combined with 5 ml of extraction buffer in duplicate Falcon Tubes to yield a 5:1 buffer:sediment ratio (*v/v*). Tubes were sonicated on ice for 60 seconds using pulse sonication (Bronson microprobe, at 20 kHz). Sediment extracts were centrifuged to remove particles from the extraction liquid (5,000 *x g*, 10 minutes, 4^{o}C) and loaded onto a Bio-Rad gel prep cell 0.5 cm diameter gel tube. The gel tubes were poured to a height of 3 cm to provide a large sample loading volume above the cast gel. THAA concentrations were used as a proxy for total protein to adjust loading volumes for gels. The gel consisted of 10% Acrylamide/Bis, 0.125 M Tris-HCl. The gel prep cell was run at 180 volts until the ion front moved approximately 1 cm down the gel. The top 1 cm was then excised for tryptic digestion.

In the slurry method, approximately 1.5 g dry weight surface sediment (~1 ml wet sediment) was combined with 1 ml of extraction buffer in duplicate Falcon Tubes to yield

a 1:1 buffer:sediment ratio (*v:v*). Tubes were sonicated on ice for 60 seconds using pulse sonication (same conditions as above) and 500 μL of sediment + extraction buffer slurry mixture was deposited onto a Bio-Rad gel prep cell 0.5 cm diameter gel tube poured to a height of 10 cm. Gel composition and running conditions for the slurry method were the same as the traditional method. The slurry gel was run until the ion front moved 5 cm down the gel. Sediment particles remained at the top of the gel and were easily washed away after the gel run was finished (Figure 3-1). The top 5 cm of the gel was then excised for digestion. Slurry mixture was also loaded onto a pre-cast 12% Bis-Tris Bio-Rad 1-dimension gel (referred to as "flat gel") and run until the ion front moved 5 cm down the gel. The top 5 cm of the slurry flat gel was then excised for tryptic digestion. The extraction buffer and sonciation process in this study were very similar to methods used to extract estuarine bacterial proteins (Kan et al., 2005) and marine particulate proteins (Dong et al., 2010) from a range of cell types. This was important to consider so that extraction was not biased towards eukaryotic cells.

Optimization of protein recovery - Once it was established that an electrophoretic extraction and preparative technique was successful, several methods for the optimization of the extraction were tested on BSA as a model protein. Deeper sediments, the 20 to 22 cm horizon of a sediment core from the Bering Sea, were utilized for all the permutations of the optimization investigation. Treatments included the type of extraction buffer and gel type used, as well as the type of preparation loaded onto the gel. Four different extraction buffers were tested: EDTA extraction buffer, $CaCl_2$ extraction buffer, SDS extraction buffer, and urea only (Table 3-1). Two types of gels were used in the extraction: preparative tube gels (run in BioRad Mini-Prep Cell) and pre-cast 12% Bis-

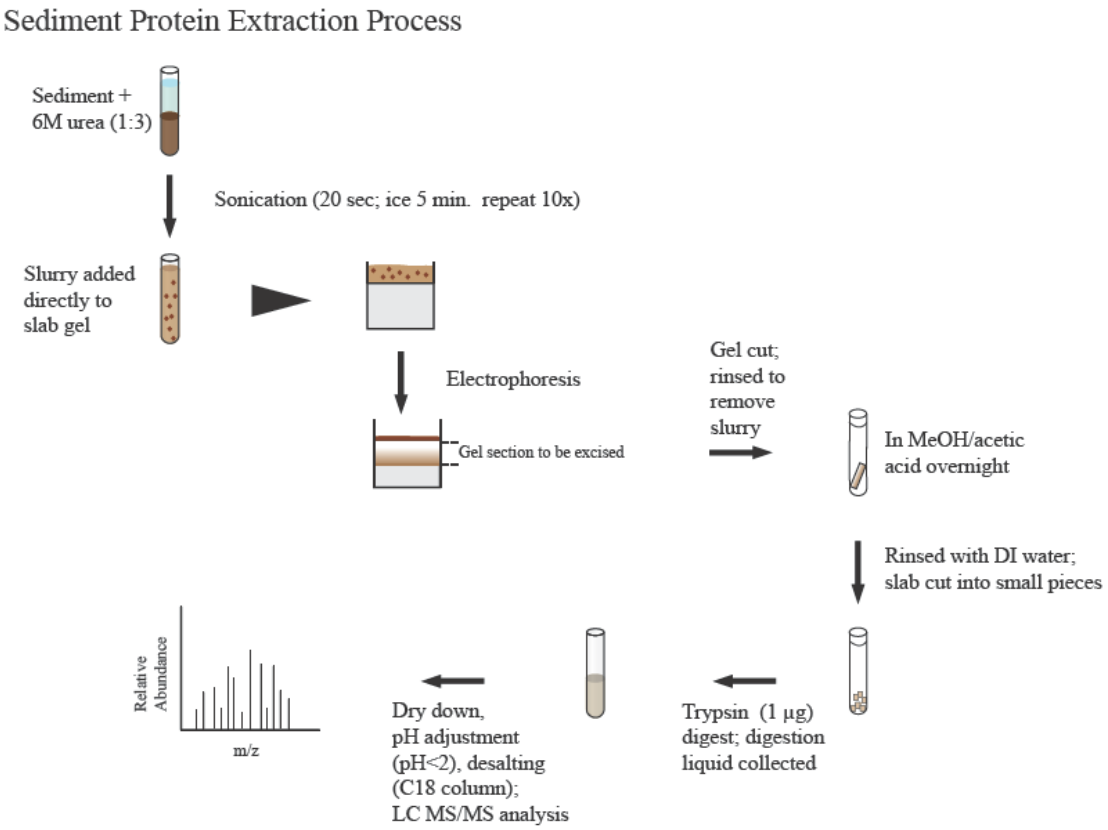

Figure 3-1. Schematic workflow for the slurry approach for extraction, purification and digestion of sedimentary samples prior to LC/MS analysis.

Tris 1-dimension, flat gels (Invitrogen NuPAGE Novex). Two types of sample preparations were tested: a traditional method where extraction buffer was loaded onto the gel and a slurry method in which a mixture of extraction buffer and sediment were loaded directly onto the gel. Fourteen permutations of gel-type, extraction buffer, and sample preparation type were investigated (see Table 3-1).

For most treatments, 750 µl of extraction buffer was mixed with 250 µl of sediment, the exception being those samples which employed only urea for extraction. To those samples 300 µl of urea was added. All extraction buffers contained the following (with the exception of the urea only samples): urea (7M), thiourea (2M), tris-HCl (0.01M), glycerol (10% v/v), ampholytes (pH - 3 to 10) (0.2% v/v) and tributylphosphine (0.002M). In addition, the EDTA extraction buffer contained: CHAPS (2% w/v) and EDTA (1mM); the $CaCl_2$ extraction buffer contained: CHAPS (2% w/v) and $CaCl_2$ (0.1 M); and the SDS extraction buffer contained: SDS (1% w/v) and EDTA (1mM). Those samples which utilized urea only were extracted with 6M urea. To all samples, 1 µg of high quality BSA was added. Samples were sonicated for 20 sec with a titanium microtip and placed on ice for 5 min. This was repeated for a total of ten sonication treatments and samples were placed in the -20°C freezer overnight. Slurry samples were thawed and the entire sample was placed on top of either the tube gel or the flat gel. Samples that did not include sediment particles were centrifuged, the supernatant was filtered, and 500 µl of extract was placed on either the tube or flat gels. Tube gels were prepared according to BioRad Mini-Prep Cell specifications for discontinuous gels (12% resolving gel with 200 µl of 12% stacking gel on top) and poured to a height of 10 cm. Tube gels were run until the ion front had migrated

Table 3-1. Combinations of extraction optimization methods tested

Extraction Buffer	Tube Gel	Flat Gel
EDTA	T,S	T,S
$CaCl_2$	T,S	T,S
SDS	T,S	T,S
Urea (only)	-	T,S

- : Protein not detected
T: Extraction buffer only
S: Extraction buffer and sediment slurry

approximately 2 cm from the top of the gel and flat gels were run until the ion front had migrated approximately 2.5 cm from the top of the gel. Gels were cut just below the ion front, rinsed three times with DI water, and placed in 50% methanol, 10% acetic acid overnight.

Trypsin digestion of SDS-PAGE slices – All samples purified via SDS-PAGE gels were digested using the same protocol. Before digestion the excised tube gels and flat gels were cut into 2 mm sized cubes to increase the exposed surface area. Pieces were covered with 100 mM ammonium bicarbonate and rinsed for 15 min followed by a 15 min rinse in acetonitrile. The rinse cycle was repeated twice more and gel pieces were dried in a vacuum drier (Speedvac) for 45 min. To the gel pieces, 1 µg of trypsin was added and the sample was placed on ice for 45 min. Reduction, alkylation, and digestion for surface sediment samples generally followed procedure by Shevchenko et al., (1996). For optimization testing, samples were removed from the ice covered with ammonium bicarbonate and digested overnight without reduction or alkylation. The pH of the sample was adjusted with 5% formic acid to a pH $\leq$ 2 and run through a C18 desalting column (Nest Group) following which samples were dried (Speedvac) and volumes were adjusted in preparation for analysis.

Direct digest of sediment – Prior to shotgun proteomic analysis, 100 mg of sediment was mixed with 300 µl 6 M urea and 50 mM ammonium bicarbonate. The sediments were then sonicated using a Bronson sonicating microprobe, at 20 kHz for 60 sec. on ice. The pH was raised by adding 18 µl 1.5 M Tris-HCl (pH 8.8). To reduce sulfhydryl linkages in proteins, 7.5 µl TCEP was added to the sediments, vortexed and incubated for 1 hr (37^{o}C). Proteins were then alkylated by adding 60 µl of 200 mM

iodoacetic acid and incubated in the dark for 1 hr. After the addition and incubation of 60 µl of dithiothreitol (1 hr room temp), the urea was diluted with the addition of 2.4 ml 25 mM ammonium bicarbonate, 600 µl HPLC-grade methanol and 1 µg of sequencing grade trypsin. The trypsin incubation was completed overnight at room temperature. Samples were centrifuged (14,000 x g, 20 minutes) and the digest with buffer removed. The sediments were then washed 3 times with 1 ml 25 mM ammonium bicarbonate, centrifuged and extracts combined. The volume was reduced to ~10 µl and 200 µl of 5% ACN, 0.1 % trifluoroacetic acid was added prior to desalting the peptides using a C18 desalting centrifuge column (NEST group). Samples were desalted using the protocol provided by manufacturer.

Mass Spectrometry – Protein analyses for all samples were conducted using standard shotgun proteomic techniques employing nanocapillary HPLC-MS/MS as described previously (Washburn et al., 2001; Aebersold and Goodlett, 2001). Samples were introduced into a hybrid linear ion Orbitrap (LTQ-OT) mass spectrometer (Thermo Fisher, San Jose, CA) via a NanoAcquity high performance liquid chromatography (HPLC) system (Waters, Beverley, MA). Trapping and analytical capillary columns were packed in-house using a pressurized cylinder (Brechbuhler AG, Schlieren Switzerland). Magic C18 (5 µm diameter, 100 Å pore size) particles (Michrom Bioresource, Auburn, CA) were slurried with analytical grade MeOH and placed in the cylinder to pack columns with 1000 psi nitrogen. Trapping column capillaries were 20 mm x 100 µm i.d., while the analytical column dimensions were 150 mm x 75 µm i.d. The trapping column was prepared with a sintered glass frit on one end and the analytical

column was tapered in a flame by gravity that allowed it to serve as a frit and electrospray ionization needle.

Chromatography was performed using acidified mobile phases: A) water, 0.1% (v/v) formic acid and B) acetonitrile, 0.1% (v/v) formic acid. Chromatography was followed as in Nunn et al. (2010). Based on parallel amino acid measurements of each sample used as a proxy of total protein, 1 μg of protein-equivalent material was injected onto the nanocapillay HPLC column for MS/MS analysis, which produced average total ion current (TIC) signal intensities of >1 x 10^7. The LTQ-OT was operated using a data-dependent acquisition (DDA) mode, where the five most intense ions from each precursor ion (MS1) scan are selected for collision induced dissociation (CID) and tandem mass spectral (MS2) detection (for review see Nunn and Timperman, 2007). Sample digests were analyzed first using a standard full scan, where the MS2 ion selection is chosen from the top five most intense ions in the *m/z* range of 350-2000. The top five most intense ions were then selected for CID by DDA from the following *m/z* ranges: *350-444, 444-583, 583-825, 825-1600* (Spahr et al., 2001; Davis et al., 2001; Yi, et al., 2002; Scherl et al., 2008).

Database Searching – The search engine SEQUEST was used to match tandem mass spectra to peptide sequences found in protein databases (Eng et al., 1994; 2008). Four protein sequence databases were evaluated for mass spectra collected from Bering Sea sediment gel digests:

1) Thaps database contains the proteome of *Thalassiosira pseudonana* (marine diatom, well annotated proteome, Armbrust et al., 2004; Oudot-Le Secq et al., 2007), and expanded with the proteomes of *Prochlorococcus marinus* (marine cyanobacterium), and

Candidatus Pelagibacter ubique (marine bacteria). These proteomes provide representation of algae, autotrophic bacteria, and heterotrophic bacteria respectively (14,795 proteins, 15 megabytes).

2) GOS/Thaps database contains the proteome of *T. pseudonana* and the Global Ocean Survey Combined Assembly Protein (GOS) database (Yooseph, et al. 2007). This concatenated database, containing protein sequences of microbes from a variety of marine environments, was used in an attempt to correlate resulting tandem mass spectra from sediments to a variety of possible bacterial proteins (6,121,580 million protein sequences, 2.3 gigabytes). Available protein names and source organisms were acquired from the CAMERA online portal (Community Cyberinfrastructure for Advanced Microbial Ecology Research & Analysis: http://camera.calit2.net/index.shtm).

3) NCBI-NR database (National Center for Biotechnology Information Reference Sequence) consists of a non-redundant collection of highly annotated DNA, RNA, and protein sequences from diverse taxa, including marine organisms (Pruitt et al., 2002). While the NCBI-NR has fewer marine protein sequences compared to the GOS database, it has greater functional information, diversity, and contains a variety of eukaryotic marine organisms not found in the GOS (11,934,213 proteins, 4.9 gigabytes);

4) NCBI-Refined database generated to include all proteins of each of the species in the list of identifications from the NCBI-NR search on the slurry surface sediment sample. This database contained the proteomes from 107 organisms and is roughly 25-fold smaller than the NCBI-NR database (417,199 proteins, 187 megabytes).

Mass spectra from the digests of the traditional, slurry, and direct digest extraction methods on Bering Sea surface sediment were searched against all four databases. Mass

spectra from the 1-D surface sediment slurry flat gel digest were only searched against the Thaps database. Two modifications were set in the SEQUEST parameter file to replicate analytical modifications completed in the lab: 57 Da fixed modification on cysteine (resulting from IAM alkylation) and 16 Da variable modification on methionine from oxidation. Predicted fragmentation vs. observed tandem mass spectra was statistically evaluated with PeptideProphet, and ProteinProphet was used to assign and group peptides into proteins (Keller et al., 2002; Nesvizhskii et al., 2003). Both PeptideProphet and ProteinProphet were set to a 90% confidence level, which corresponds to a predicted 10% error rate. Ignoring these confidence limits and/or false discovery rates will yield inaccurate and spurious protein identifications. To quantify model protein recoveries, a fifth specific database containing BSA and 50 common contaminants was used (51 proteins, 32.9 kilobytes). Spectra generated from the extraction optimization investigations were searched against this database with sequence coverage (%) of BSA used to determine the most successful extraction procedures. As stated earlier, for all database evaluations only proteins reported with high confidence (>90%) were accepted and discussed in this study.

Amino Acid Analysis – To compare amino acid composition of initial sediment and the treatments, individual amino acids were identified and quantified by gas chromatography mass spectrometry (GC-MS) using the EZFaast method (Phenomenex ®). All samples were dried and hydrolyzed for 4 hours at 110^{o}C with analytical biology grade 6 M HCl (Cowie and Hedges, 1992b), and L-γ-Methylleucine as the recovery standard (Waldhier et al., 2010). Following hydrolysis and derivatization, amino acids were quantified using an Agilent 6890 GC with samples injected at 250^{o}C and separated

through a DB-5MS (0.25 mm ID, 30 m) GC column with hydrogen as the carrier gas. For amino acid identification the GC was coupled to an Agilent 5973 mass spectrometer run under the same conditions. Helium was used as the carrier gas for the amino acid analysis in the GC, and acquisition of spectra between 50-600 Da mass range were collected. Bovine serum albumin (BSA) was analyzed in parallel to correct for responses among individual amino acids and calculation of molar ratios.

Assessment

Sediment properties and amino acids – The surface sediment examined in this study contained 0.48% organic carbon (OC) and 0.06% particulate nitrogen (PN) while the deeper sediments utilized in the optimization experiments contained 1.07% OC and 0.15% PN. Amino acids represented 1.39 $\pm$ 0.12 mg THAAs/g sediment dry weight in surface sediments, similar to other northern latitude marine sediments (Mintrup and Duinker, 1994; Horsfall and Wolff, 1997). Deep sediments accounted for 2.59 $\pm$ 0.67 mg THAA/g sediment dry weight. THAAs contributed 28.9% of POC and THAA-N contributed 29.1% of PN for surface sediments and 24.4% of POC and 31.2% of PN for deep sediments. Extraction efficiency was based on THAA recovery and was calculated for surface sediment samples resulting in efficiencies of 12.5% $\pm$ 1.1 for buffer surface sediment extraction, and 100.6% $\pm$ 4.5 for buffer surface sediment slurry mixture compared to whole sediment. Amino acid distributions in surface sediments showed only subtle differences between whole surface sediment and the two surface sediment extraction methods (Figure 3-2).

Database and method evaluation of identified proteins from surface sediments – The search against the Thaps database of the slurry method resulted in the greatest number of confident protein identifications. Using the database that contained proteomes from one diatom and two marine bacteria, 302 unique peptides were identified from the slurry tube gel method, which correlated to 126 protein identifications (Table 3-2, Appendix 3-1). The slurry 1-D flat gel method identified 31 proteins (82 peptides). The traditional method and the direct digest retrieved 60 proteins (149 peptides) and 6 proteins (7 peptides), respectively. The majority of proteins identified from each of the

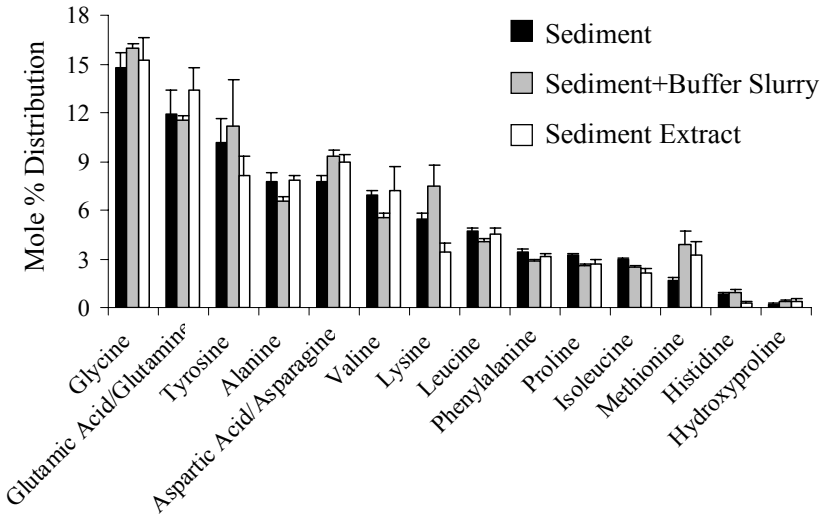

Figure 3-2. The comparative distribution of amino acids observed in Bering Sea shelf sediments using the two extraction approaches verses hydrolysed intact sediments. Order is based on average mole % values for each amino acid. Error bars represent standard deviation of each measurement.

Table 3-2. Total proteins identified by extraction method and proteomic database. Proteins = number of proteins identified; Peptides = number of peptides identified; *T. pseudo* = number of *Thalassiosira pseudonana* proteins identified; *T. pseudo+* = number of proteins identified conserved among *Thalassiosira pseudonana* and another source.

Database	Extraction Method	Proteins	Peptides	*T. pseudo*	*T. pseudo+*
Thaps	Direct Digest	6	7	6	0
	Slurry Flat Gel	31	82	30	1
	Traditional Tube Gel	60	149	60	0
	Slurry Tube Gel	126	302	122	0
GOS	Direct Digest	4	4	0	0
	Traditional Tube Gel	63	130	37	10
	Slurry Tube Gel	114	257	87	7
NCBI-NR	Direct Digest	16	16	2	0
	Traditional Tube Gel	31	115	15	7
	Slurry Tube Gel	44	115	16	16
NCBI-Refined	Slurry Tube Gel	84	205	37	44

four methods correlated to identifications from the diatom, *T. pseudonana*. Only two *P. marinus* protein identifications from the slurry tube gel method were made with no *P. marinus* identifications in the traditional gel or direct digest. There were no proteins identified as *C. P. ubique* using any of the extraction methods. Among the three methods the slurry gel and traditional gel methods had 46 protein identifications in common (Figure 3-3A). The slurry tube gel, traditional gel, and direct digest methods had five protein identifications in common.

Proteins identified from the Thaps database were grouped by Gene Ontology functional categories (Table 3-3) to compare the distributions between different extraction techniques. The majority of proteins identified were involved in metabolic processes. The biggest difference between the distribution of metabolism proteins between the slurry tube gel and traditional tube gel methods was the large contribution of translation/ribosomal proteins among slurry tube gel identifications at 20.2% (26 protein identifications), versus the small contribution from the same category identified from the traditional tube gel method at 1.7% (1 protein identification). The only unique protein identified by the direct digest method was chloroplast ribose-5-phosphate isomerase, a phosphate shunt protein.

Tandem mass spectra searched against the GOS/Thaps database yielded 114 protein identifications from the slurry tube gel method, 63 from the traditional tube gel method, and 4 from the direct digest method (Table 3-2, Appendix 3-1). There were fewer peptides identified using each extraction method for the GOS search compared to the Thaps search. The majority of proteins identified from the slurry tube gel method still primarily correlated with sequences associated with the diatom proteome, even though

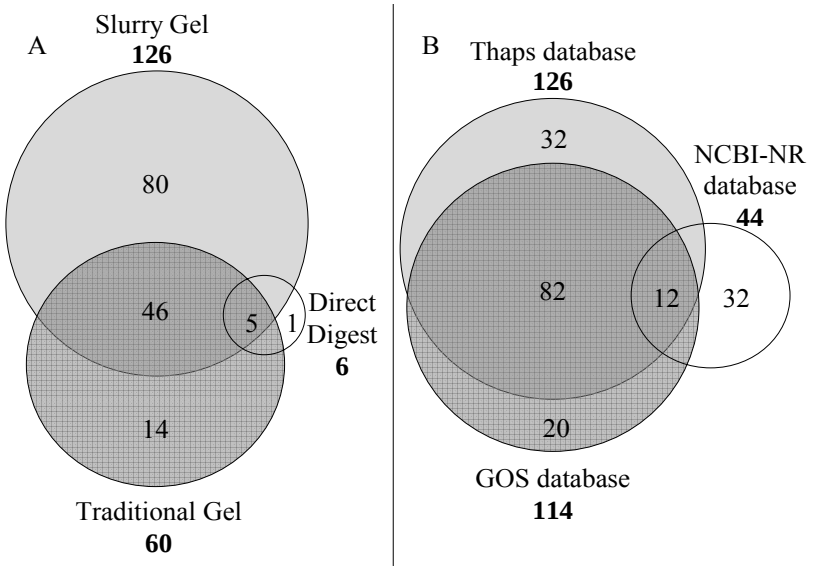

Figure 3-3. Venn diagrams of (A) the number of proteins identified in common between the slurry tube gel, traditional tube gel, and direct digest methods searched against the Thaps database; (B) number of proteins in common between the Thaps, GOS/Thaps, and NCBI-NR database searches of the slurry tube gel method.

Table 3-3. The cellular functions of identified proteins found using the Thaps database organized as subgroups of function (Gene Ontology Consortium, 2008). The numbers indicate the number of identifications with numbers in parentheses as the percentages of total proteins found using each method.

Function	Slurry Tube Gel	Standard Tube Gel	Direct Digest
Metabolism	96 (77.4%)	41 (68.3%)	5 (83.3%)
Photosynthesis	36 (29.0%)	23 (38.3%)	4 (66.7%)
Translation, Transcription	26 (21.0%)	1 (1.7%)	-
Metabolism, Recycling	8 (6.5%)	5 (8.3%)	-
Glycolysis, Respiration	8 (6.5%)	4 (6.7%)	-
Enzyme	7 (5.6%)	5 (8.3%)	-
Biosynthesis	4 (3.2%)	1 (1.7%)	-
GTPase	4 (3.2%)	-	-
Modification	3 (2.4%)	2 (3.3%)	-
Pentose-Phosphate Shunt	-	-	1 (16.7%)
Binding, Structure	14 (11.3%)	8 (13.3%)	-
Binding DNA, RNA	4 (3.2%)	-	-
Binding ATP, GTP	3 (2.4%)	-	-
Heat Shock	3 (2.4%)	3 (5.0%)	-
Structure	2 (1.6%)	-	-
Folding	1 (0.8%)	-	-
Binding, Zn	1 (0.8%)	-	-
Binding, Protein	-	5 (8.3%)	-
Transport	11 (8.9%)	7 (11.7%)	1 (16.7%)
Transport, Proton	5 (4.0%)	5 (8.3%)	1 (16.7%)
Transferase	2 (1.6%)	-	-
Transport	2 (1.6%)	2 (3.3%)	-
Transport, Protein	1 (0.8%)	-	-
Nucleotidyltransferase Activity	1 (0.8%)	-	-

- : Protein not detected

the database consisted almost exclusively of marine bacterial proteins, as these proteins represented higher confidence identifications. Only 20 proteins correlated uniquely to GOS microbial protein sequences (Figure 3-3B), consisting mostly of metabolism proteins and fewer transport and binding proteins. Seven proteins were identified with peptide sequences that were identical between diatom and marine bacterial proteins. Thirty seven of the traditional tube gel proteins correlated to *T. pseudonana*, 16 from GOS, and 10 as both *T. pseudonana* and GOS proteins (i.e. homologous sequences).

Mass spectral results searched against the NCBI non-redundant (NR) database yielded similar distributions, with the most proteins being identified from the slurry tube gel method and the least from the direct digest (Table 3-2). Fewer peptides and proteins were identified using the NR database compared to the Thaps and GOS databases. Proteins identified as originating from diatoms or conserved among diatoms and other organisms made up the majority of proteins identified from the slurry and traditional tube gel methods (Appendix 3-1). The majority of direct digest identified proteins were bacterial in origin. The NCBI-Refined database search yielded more peptide and protein identifications than the full NCBI-NR database search for all extraction methods. Two proteins from the full NCBI-NR and NCBI-Refined searches, separate ABC transporter proteins, were uniquely identified as originating from prokaryotic sources.

Protein recovery optimization from sediments – There was successful recovery of BSA standard from sediment samples with most extraction methods tested (Table 3-4). The exceptions to this were the two $CaCl_2$/tube gel combinations in which no identifiable BSA was recovered from either the traditional or slurry methods. Sequence coverage, used as a metric to determine the most efficient extraction methods, ranged from 0% (no

BSA recovered) to 22% (Table 3-4). Both samples extracted with the urea extraction buffer resulted in the recovery of the highest number of independent spectra, unique peptides, and sequence coverage of BSA. The slurry method yielded 22% sequence coverage while the traditional method returned 13%.

Table 3-4. Results of protein search for BSA (Bovine Serum Albumin) in extraction optimization experiments. Total independent spectra indicates the number of mass spectra that were correlated with BSA peptides.

Gel Type	Sequence coverage (%)	Number of unique peptides	Total independent spectra
Flat			
EDTA, traditional	9.9	4	36
EDTA, slurry	5.6	2	5
$CaCl_2$, traditional	7.4	3	29
$CaCl_2$, slurry	10.4	4	33
SDS, traditional	4.6	2	3
SDS, slurry	8.1	6	30
Urea, traditional	13.2	8	46
Urea, slurry	22.1	13	45
Tube			
EDTA, traditional	3.3	2	2
EDTA, slurry	9.1	4	4
$CaCl_2$, traditional	-	-	-
$CaCl_2$, slurry	-	-	-
SDS, traditional	9.6	3	3
SDS, slurry	5.9	2	2

- : Protein not detected
Traditional: extraction buffer only added to gel
Slurry: extraction buffer and sediment added to gel

Discussion

The successful identification of a variety of proteins and the ability to recover standard protein addition using the slurry and traditional gel methods demonstrate that electrophoresis provides an effective isolation method for proteins in sediment systems. The greater number of peptides and proteins identified using the slurry gel method compared to the traditional gel method demonstrates that the electric field applied directly to sediment particles can enhance protein extraction. Along with the slurry method, the Thaps database proved to be the most effective database at maximizing protein identifications for the Bering Sea system. Rather than using gel electrophoresis as a means for visualizing the isolated proteins, we employed the SDS-PAGE technique to enhance protein solubilization from the sediment matrix and as a stabilizing matrix to remove contaminants. The excised gel can then undergo a standard rinse process followed by in-gel protein digestion using trypsin and elution of the resulting peptides for tandem mass spectrometry analysis. The low numbers of confident protein identifications using the direct digest of the sediments suggests that the suite of products, or perhaps the solid matrix itself, interfere with protein digestion and identification.

As the availability of proteins for analysis in marine sediments may be limited by degradation or the binding of proteins to various matrices, it is beneficial to optimize the extraction technique in order to maximize the recovery of any available protein material. Chen et al. (2008) demonstrated the applicability of sequence coverage as a measure of protein expression between subjects. Though the utilization of sequence coverage as a measurement tool demonstrated greater intra-sample variability than other methods tested (Chen et al., 2008) it is an easily determined variable and its use is acceptable as a metric

in the optimization experiments as a determination of the effectiveness of the various protein extraction methods.

Comprehensive testing of BSA with the extraction buffers, electrophoresis gels and sample preparations found extraction with 6M urea and the placement of the slurry directly on a flat gel most effective for BSA sequence coverage recovery (Figure 3-1). The slurry method also yielded 9% greater sequence coverage of BSA than the more traditional method. This supports the observation that the application of an electric field to the slurry mixture enhanced qualitative protein extraction.

The comparison of the Thaps, GOS/Thaps, NCBI-NR, and NCBI-Refined databases sheds some light on the amount of information that can be gained on protein functions and taxonomy of source organism at different levels (i.e. kingdom, class, family) from complex samples. In addition, we can evaluate the usefulness of large databases as it relates to search-time requirements and available computational resources. The distinct advantage of using the more complex GOS and NCBI-NR protein databases are the greater number of organisms from which proteins can be identified. The disadvantage of searching these databases is the reduced statistical significance and sensitivity, and the amount of computational resources and dedicated time involved. Analysis of 10 tandem MS files (each file containing thousands of spectra) using the GOS database consumed >720 hours, and the NCBI-NR database >1080 hours, using an 800 CPU cluster. This is a large amount of time compared to the Thaps database searches of <5 hours. Because these searches consumed so much computational time, we focused this study on the search results from only the tandem MS analyses performed on the slurry tube gel sample treatment as it revealed the greatest number of high confidence

multiple-peptide protein identifications with 67, and 59 single peptide protein identifications.

Each database used, Thaps, GOS/Thaps and NCBI-NR, contained the entire *Thalassiosira pseudonana* proteome. Using the GOS/Thaps database, only 20 of the 114 proteins identified were not *T. pseudonana* in origin (~15%). Searches against the NCBI-NR database identified 12 proteins originating from organisms other than *T. pseudonana*. Of the 12 non-*T. pseudonana* proteins identified using the NCBI-NR database, 9 of them still correlated to different marine diatoms sources than *T. pseudonana*. Despite some differences in identified protein source resulting from the use of protein sequence databases with proteins from different species, assigned protein functions were the same for over 95% of peptides identified from multiple databases (Figure 3-4). Although we searched the same suite of tandem MS data against different databases, the larger databases (e.g. GOS and the NCBI-NR) yielded fewer confident protein identifications, precisely because of the fact that they include an additional 6 to 11 million proteins that are not from *T. pseudonana*. These results demonstrate that few novel non-diatom protein identifications were made with the large databases searched against Bering Sea surface sediment extract digests, and that of identifiable proteins from non-algal sources were not being ignored.

Interestingly, many of the peptides that were identified to be from *T. pseudonana* when searched against the Thaps database were not confidently identified using either of the other two databases. This results from the inability of PeptideProphet to decipher and report homologous peptides with confidence. Peptide prophet uses a correlation of two parameters: 1) how well the observed spectra matches the theoretical spectra (xcorr); 2)

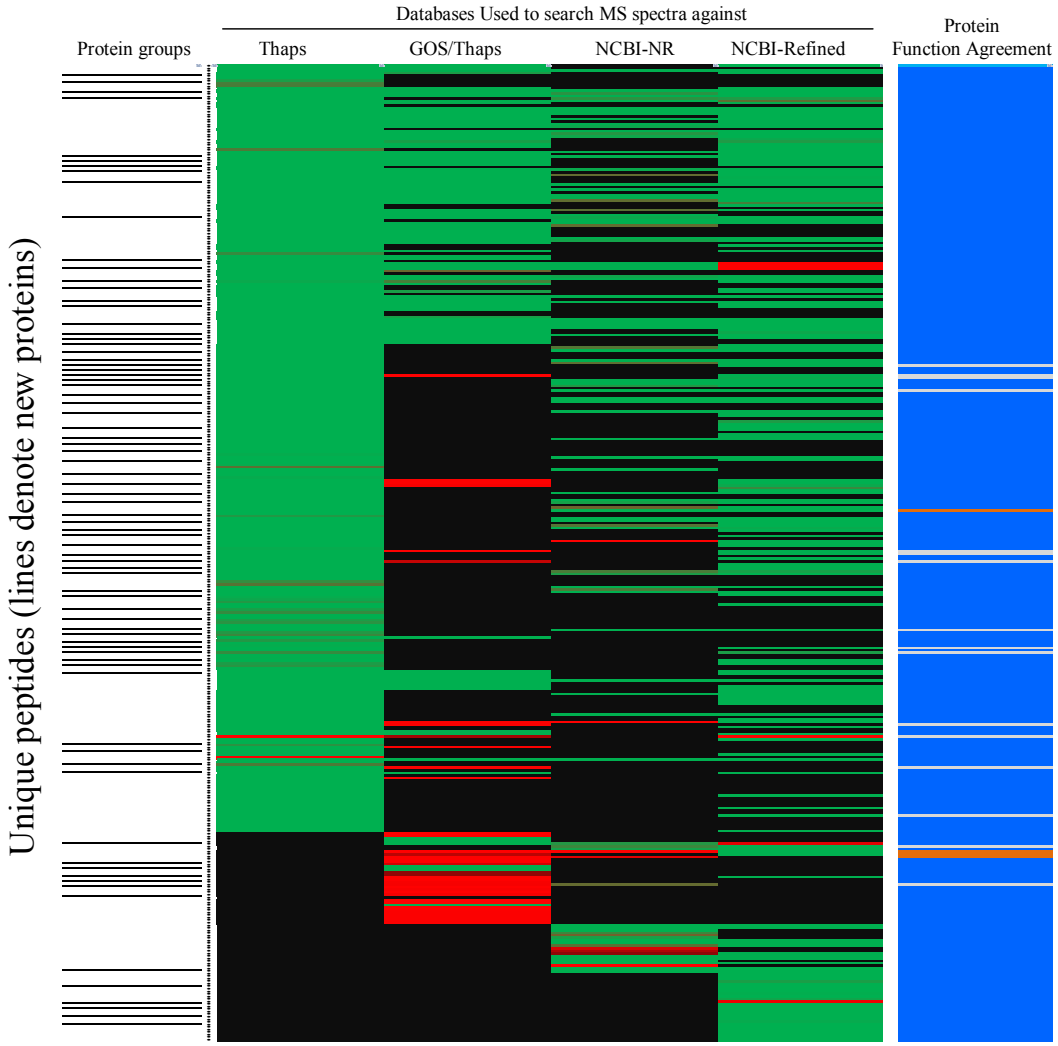

Figure 3-4. Species assignment and protein function assignment comparison of results from SEQUEST search of surface sediment slurry tube gel digest mass spectra using 4 different databases: Thaps (column 1); GOS/Thaps (column 2); NCBI-NR (column 3); NCBI-Refined (column 4). 385 unique peptides are represented along the vertical axis and represented as dots. Peptides are grouped together from the same protein and the black lines on the far left mark the beginning of a new group of peptides associated with the same protein. Species assignments are represented by shades of green or red and black. Green indicates the peptide was designated to originate from a marine or aquatic eukaryote. Red indicates the peptide was designated to originate from a marine, aquatic, or soil bacteria. For the case of greens and reds, the brighter the color, the higher confidence peptide prophet gave the assignment (e.g. $p > 0.99$), whereas lighter shades of red or green indicate poorer peptide correlations ($0.9 < p < 0.99$) Black indicates that the peptide was not assigned to a protein using that particular database. The far right column is a color-coded illustrate if the function of the protein assigned is in agreement between the 4 databases searched. Blue indicates function is the same, grey indicates function is unknown by one or more databases, orange indicates function does not agree between database assignments. Functional agreement is present in over 95% of the peptides.

how different the first peptide match is from the second peptide match (ΔCorr). In general, an assignment is made if the xcorr > 2.0 and ΔCorr > 0.1. When using a larger database there is more peptide sequence similarity (e.g. SEVSALLGR, SEVSAILGR). As a result, PeptideProphet will assign a low ΔCorr to the second best peptide match. A low ΔCorr will decrease the overall statistical confidence and SEQUEST will not report any peptide match, even at high xcorr values.

The greatest number of identical peptide and protein assignments from different database searches was observed when searching the Thaps and GOS/Thaps databases (Figure 3-3B). As mentioned earlier, all databases included *T. pseudonana* and the protein assignments that were identical between the GOS/Thaps and Thaps database searches were all *T. pseudonana* in origin. Despite fewer identified proteins, the larger databases do provide breadth to the sources of conserved proteins. Peptides from several identified proteins were conserved among *T. pseudonana* and hundreds of other organisms. Given the context of the system, seasonally diatom dominated Bering Sea, and that other peptides are predominantly identified uniquely to *T. pseudonana* using all three databases, it is likely that these conserved proteins are also diatom in origin.

Given that the Thaps database search of the slurry method yielded the greatest number of identifications, it not surprising that the Thaps search against the slurry method data also identified a suite of proteins with the greatest range of isoelectric points (pI). Of Thaps identified proteins, a total of 43 slurry tube gel proteins and only 5 traditional tube gel proteins were identified with an isoelectric point above eight. The isoelectric point of a molecule is the pH at which the molecule carries no net charge, and is also the point at which the solubility of the molecule is at its lowest. The much larger

proportion of high pI proteins identified illustrates the greater electrophoretic extraction achieved by the slurry tube gel method.

Basic proteins with high pI carry a more negative charge would likely be more tightly bound to positively charged functional groups in sediments (Henrichs, 1993). The extraction buffer used in this study, containing high concentrations of protein solubilizing reagents urea, thiourea, CHAPS, EDTA, is slightly acidic and thus not as effective at extracting basic proteins as electrophoretically assisted extraction. Of the 37 slurry tube gel identified proteins with a pI > 9, 20 are structural constituents of ribosomes. No ribosome structural constituents were identified using the traditional tube gel technique. This shows that the slurry tube gel method not only extracts a greater number of proteins than the traditional tube gel method, but a wider range of protein functionalities as well.

Surface adsorption has been described as an important interaction between organic matter and sedimentary minerals (Mayer, 1994; 1999; Mayer et al., 2002). Various mechanisms may be involved in the adsorption of organic matter to mineral surfaces including van der Waals interactions (Rashid et al., 1972), ligand exchange (Davis, 1982), cation bridges (Greenland, 1971), cation exchange (Wang and Lee, 1993), anion exchange (Greenland, 1971), and hydrophobic effects (Nguyen and Harvey, 2001). The mechanisms of interaction described above between protein, sedimentary minerals, and organics often include some form of charge interaction. The electric field of gel electrophoresis may disrupt these interactions allowing intact protein material to mobilize into the gel. The electrokinetic phenomenon was first observed by Reuss (1807) when the application of a constant electric field caused migration of aqueous clay particles in water. Fractionation of mineral species by electrophoresis was later demonstrated by

Dunning et al. (1982). This principle of mineral mobilization by an electric field may be central in the liberation of proteins from sediments.

Recommendations for methodologies

This is the first study to use gel electrophoresis as an extraction method for the recovery of protein from sedimentary matrices. Initial results, further supported by optimization evaluations, allowed us to hypothesize that electric current disrupts the interactions between protein and sediment in order to mobilize protein into the electrophoresis gel. Optimization experiments with BSA show that the most effective extraction of peptides from sediments occurs through the use of a urea extraction buffer, pre-cast 1-D flat gel and the application of the sediment and buffer combination directly to the gel. By adding the extraction buffer with the sediment to the gel, proteins are solubilized and removed more efficiently from the particles, while the gel is an excellent trapping matrix for the proteins so that contaminants can be adequately washed away prior to enzymatic digestion and MS analysis. This study demonstrates that complex protein databases, while providing more potential protein sources, do not necessarily translate into a greater number or more confident protein identifications. Fewer protein identifications with larger protein databases appear to be due to statistical issues with PeptideProphet. Furthermore, functional-level information is retained, despite the organism with which the protein is associated due to sequence homology. This finding demonstrates that identifying proteins from mixed (often unknown) communities can be accomplished at the protein function-level, although determining and/or targeting the specific species the protein originated from remains difficult. The large contribution of diatom identified proteins in Bering Sea sediment from the simple Thaps database and the complex GOS and NCBI-NR databases indicates that primary production is an important source of protein material to continental shelf sediment in this system.

Chapter 4: Identifying and tracking proteins through the marine water column: insights into the inputs and preservation mechanisms of protein in sediments

Abstract

Proteins generated during primary production represent an important fraction of marine organic nitrogen and carbon, and have the potential to provide organism-specific information in the environment. The Bering Sea is a highly productive system dominated by seasonal blooms and was used as a model system for algal proteins to be tracked through the water column and incorporated into detrital sedimentary material. Samples of suspended and sinking particles were collected at multiple depths along with surface sediments on the continental shelf and deeper basin of the Bering Sea. Modified standard proteomic preparations were used in conjunction with high pressure liquid chromatography-tandem mass spectrometry to identify the suite of proteins present and monitor changes in their distribution. In surface waters 207 proteins were identified, decreasing through the water column to 52 proteins identified in post-bloom shelf surface sediments and 24 proteins in deeper (3490 m) basin sediments. The vast majority of identified proteins in all samples were diatom in origin, reflecting their dominant contribution of biomass during the spring bloom. Identified proteins were predominantly from metabolic, binding/structural, and transport-related protein groups. Significant linear correlations were observed between the number of proteins identified and the concentration of total hydrolysable amino acids normalized to carbon and nitrogen. Organelle-bound, transmembrane, photosynthetic, and other proteins involved in light harvesting were preferentially retained during recycling. These findings suggest that

*Eli K. Moore, Brook L. Nunn, David R. Goodlett, and H. Rodger Harvey. Identifying and tracking proteins through the marine water column: insights into the inputs and preservation mechanisms of protein in sediments. *Geochimica et Cosmochimica Acta*, accepted.

organelle and membrane protection represent important mechanisms that enhance the preservation of protein during transport and incorporation into sediments.

Introduction

As the building blocks of proteins, amino acids represent the largest portion of characterized biochemicals in most marine environments, and are important contributors to both carbon and nitrogen pools (Burdige and Martens, 1988; Hedges, 1991; Benner et al., 1992; Lee et al., 2000). Estimates suggest that amino acids make up 2-30% of organic carbon and 15-42% of organic nitrogen in coastal and deep ocean sediments (Wakeham et al., 1997; Keil, 1999). Evidence from solid state N-15 NMR spectroscopy shows that the majority of organic nitrogen present in dissolved and particulate marine organic matter contains amide bonds, like those that occur in proteins (McCarthy et al., 1997; Knicker, 2000; Zang et al., 2001). In several cases, protein products as well as intact proteins have been observed in deep ocean waters (Tanoue, 1992; Suzuki et al., 1997; Dong et al., 2010).

Traditional approaches of measuring hydrolysable amino acids (THAAs) in environmental samples as a proxy for total protein material destroy the amide bond, and thus the inherent source information encoded in the protein's amino acid sequence is lost. Our aim is to link the observed THAA distributions to the defined amino acid sequences of identifiable proteins. This will enhance the information obtained between the biosynthetic building blocks of macromolecular organic nitrogen and the geochemical fate of their protein products. The recent characterization of the marine diatom *Thalassiosira pseudonana* genome (Armbrust et al., 2004; Oudot-Le Secq et al., 2007) and now the proteome (Nunn et al., 2009) provides insight into the biochemical pathways utilized by marine diatoms. In addition, it presents the opportunity to track the distribution of proteins from this major algal contributor in natural settings.

With primary production rates up to 570 g C m^{-2} y^{-1}, the Bering Sea is perhaps the most productive region in the world (Sambrotto et al., 1986; McRoy et al., 1987; Walsh et al., 1989). As in many high latitude systems, diatoms dominate spring production (Banahan and Goering, 1986; Springer et al., 1996), leading to a carbon export flux in the range of 10 mmol C m^{-2} day^{-1} (Chen et al., 2003). This combination of high productivity of known algal communities and rapid transport to sediments make the Bering Sea an ideal system to study the early diagenetic fate of algal proteins in marine systems.

Here we link proteomic approaches with geochemical cycling to examine the environmental fate of proteins in a system where diatoms provide the bulk of new organic matter. We have merged traditional protein buffer extraction and gel electrophoresis purification techniques with mass spectrometry-based proteomics in order to identify individual proteins in these complex environmental samples. The goal of this study was to track proteins derived from the spring diatom bloom in the Bering Sea through the water column to eventual incorporation into the sedimentary organic nitrogen pool. In doing so, we have the opportunity to identify potential mechanism(s) which regulate the distributions observed during the initial stages of diagenesis.

Methods

Bering Sea sample collection – All suspended particles from filtration, particle trap, and sediment samples were collected on the Bering Sea outer shelf and basin during the Bering Sea Ecosystem Study (BEST) cruises in the spring and summer of 2009 (Fig. 4-1). Particulate and trap samples were collected at multiple depths during the spring diatom bloom as ice retreated; samples included both suspended and sinking material inclusive of chlorophyll maximum to bottom waters (Table 4-1). Suspended particles, collected by Niskin bottles, were filtered onto 47 mm combusted glass fiber filters (GF/Fs) at three depths: chlorophyll max (4 m), 50 m, and 100 m. Sinking particles were collected from 12 hour trap deployments (40 m, 60 m, and 100 m) at the same location as suspended particles. No preservatives were used with particle trap cups prefilled with brine solution prior to deployment. After collection, 100 ml aliquots of the particle trap samples were filtered onto combusted 25 mm GF/Fs. Undisturbed sediments were collected using a multicore on both the outer shelf and in the deeper basin (3490 m) before the spring phytoplankton bloom and two months after the bloom to allow sampling of recently arrived material at the sediment-water interface (Table 4-1). Visible phytodetritus was present post-bloom on the surface of both shelf and basin sediments. Surface material (0-1 cm) was removed and samples were frozen and stored at -70^{o}C until analysis.

Amino acid analysis – To provide a metric for comparison of protein content and more traditional measures, total hydrolysable amino acids (THAAs) were quantified and analyzed in parallel with protein identification. Individual amino acids were identified and quantified by gas chromatography/mass spectrometry (GC/MS) using the EZFaast

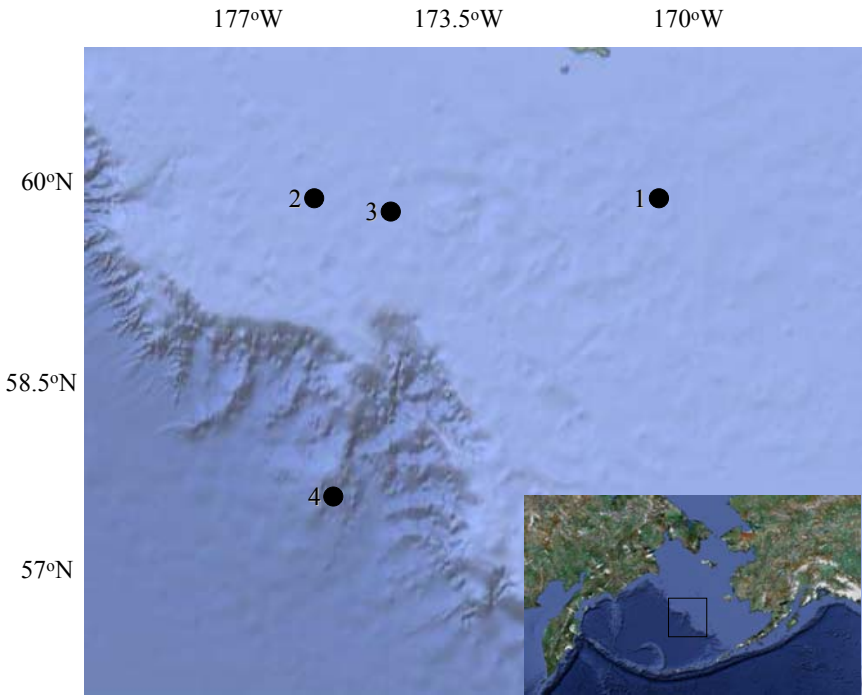

Figure 4-1. Map of Bering Sea and sample locations. Samples include: (1) shelf surface sediment during ice cover prior to the spring phytoplankton bloom; (2) water column suspended particles and sinking particle trap material during the spring phytoplankton bloom; (3) shelf sediments subsequent to the spring phytoplankton bloom and (4) basin surface sediments.

Table 4-1. Station locations and bulk properties of suspended particles (POC), trap material and sediments of Bering Sea samples. Carbon and total hydrolysable amino acid (THAA) concentrations are μg/l for suspended particles, μg/hr for trap material, and μg/g for sediments. PBS = Post-bloom Shelf; PBB = Post-bloom Basin; OWS = Over-wintered Shelf.

Sample	Date	Lat (N)	Long (W)	Depth (m)	°C	C (μg/l)	C:N Ratio	THAA (μg)
Chl Max POC	4/30/2009	59.9	176.1	4	-0.79	1172	6.2	520.0
50m POC	4/30/2009	59.9	176.1	50	-0.05	114.4	3.5	14.2
100m POC	4/30/2009	59.9	176.1	100	0.71	126.1	3.9	9.0
40m Trap	4/30/2009	59.9	176.1	40	-0.76	225	4.1	4.6
60m Trap	4/30/2009	59.9	176.1	60	0.45	314	4.0	4.0
100m Trap	4/30/2009	59.9	176.1	100	1.00	365	8.1	2.9
PBS	7/5/2009	59.6	175.2	136	1.08	10200	7.9	846.9
PBB	6/26/2009	57.5	175.2	3490	3.67	11600	7.5	687.8
OWS	4/9/2009	59.9	171.6	101	-0.15	10500	8.2	725.4

method (Phenomonex ®) which uses derivatization of AAs with propyl chloroformate and propanol for sensitive detection (see Waldhier et al., 2010 for a comparison of methods). Briefly, suspended particles, particle traps, and sediment samples were hydrolyzed for 4 hours at 110 ^{o}C (Cheng et al., 1975; Cowie and Hedges, 1992) with 6 M analytical-grade HCl and L-γ-Methylleucine as the recovery standard. Following hydrolysis and derivatization, amino acids were quantified using an Agilent 6890 capillary GC with samples injected at 250 ^{o}C and separated via a DB-5MS (0.25 mm ID, 30 m) column with H_2 as the carrier gas. The oven was ramped from an initial temperature of 110 ^{o}C to 280 ^{o}C at 10 ^{o}C per minute followed by a 5 minute hold. For amino acid identification, the GC was coupled to an Agilent 5973N mass spectrometer run under the same conditions with helium as the carrier gas and acquisition of spectra over the 50-600 Da range. Bovine serum albumin (BSA) was analyzed in parallel to correct for responses among individual amino acids and calculation of molar ratios. Amino acids were normalized to percent carbon or nitrogen using bulk samples analyzed by standard combustion methods.

Protein extraction – To extract proteins from suspended particle and particle trap samples, filters were sliced into sections, fully submerged in 3 ml of extraction buffer, (7 M urea, 2 M thiourea, 0.01 M Tris-HCl, 1 mM EDTA, 10% v/v glycerol, 2% w/v CHAPS, 0.2% w/v ampholytes, 2 mM Tributyl phosphine, see Kan et al., 2005) and pulse sonicated on ice with a Branson 250 Sonifier sonication probe at 20 kHz for 1 minute. The sonication probe was in full contact with filters and particulate material to enhance extraction. The extraction process of sonication in concentrated urea denaturing solution was very similar to past studies on estuarine bacteria (Kane et al., 2005) and marine

particulate proteins (Dong et al., 2010) identified from multiple cell types by proteomic mass spectrometry. Extracted material was then centrifuged for 5 min (5,000 *x g*) to remove particles. For suspended and sinking particles, protein extraction efficiencies were estimated by comparing Bradford Assay protein concentrations of protein extracts to total hydrolysable amino acids (THAAs) of whole samples as a proxy for total protein. The extraction efficiencies for chlorophyll max, 50 m, and 100 m suspended particulate samples (Bradford Assay Protein Concentration/THAA concentration) were 105%, 81%, and 90% respectively.

For sediment protein extraction, known weights of sediment were treated with 500 μl of extraction buffer in 1.5 ml Eppendorf tubes and pulse sonicated for one minute on ice. The entire sediment+extract buffer mixture was then loaded onto gel-prep cell tubes for isolation and molecular weight class separation of proteins. Amounts of protein material loaded onto gels were determined by measuring the concentration of THAAs in filter extracts and sediment buffer mixtures as a proxy for total protein. The extraction efficiencies (THAA concentration of extract/THAA concentration of sediment) were consistent, ranging from 11-12% for shelf and basin sediments. The extraction protocol and efficiency is described in greater detail in Moore et al (in review).

To purify extracts and separate proteins based on molecular weight (MW), 1 ml of extract from each sample was loaded onto individual preparatory electrophoresis tubes (prep-gel: 17% Acrylamide/Bis, 0.125 Tris-HCl) for one dimensional separation. Gels were covered with running buffer (0.25 M Tris, 0.192 M glycine, 0.1% SDS, pH 8.3; 180 V) and run until the ion front traveled 7 cm from the top of the gel. After electrophoresis, gels were washed and cut into five molecular weight ranges (<10, 10-25, 25-50, 50-100,

and >100 kDa) based on prior separations of MW standards (Kaleidoscope) using identical gels.

In-gel protein digestion – Before enzymatic digestion, gel MW sections were cut into 2 x 2 mm slices to increase surface area for enzyme and chemical access. Pieces were covered with 100 mM ammonium bicarbonate and rinsed for 15 min to hydrate gel sections followed by 15 min rinse in acetonitrile to dehydrate gel sections and remove detergents and other chemical interferences. The rinse cycle was repeated five times and gel sections then were dried by speed-vac for 45 min. Subsequent reduction, alkylation, and digestion followed standard procedure by Shevchenko et al. (1996). Digests were dried and volumes were adjusted to give a final protein concentration of 1 μg protein/10 μl based upon THAA concentrations and measured recoveries.

Mass spectrometry and database searching – Proteins were identified via shotgun proteomics with samples introduced into the ion trap (LTQ Velos) mass spectrometer (Thermo Fisher) via NanoAcquity high performance liquid chromatography (HPLC, Waters) (Nunn et al., 2010). New analytical and trapping columns were packed in-house prior to batch analyses of Bering Sea samples in order to ensure no proteins were carried over from previous cell lysate proteomic experiments. Analytical columns were made using 11 cm long, 75 μm i.d. fused silica capillaries packed with C18 particles (Magic C18AQ, 100 Å, 5 μm; Michrom, Bioresources) preceded by a 2 cm long, 100 μm i.d. trapping-column (Magic C18AQ, 200 Å, 5 μm; Michrom). Samples were loaded onto the trapping column with a flow rate of 4 μl min^{-1} (7 min), and then entered the analytical column at a flow rate of 250 nl min^{-1} (total run time 100 min). Peptides were eluted using an acidified (formic acid, 0.1% v/v) water-acetonitrile linear gradient (5 to 35%

acetonitrile in 60 min), and ionized in atmospheric pressure before entering the mass spectrometer. Following a survey of the ions that entered the ion trap (MS^1), the fourteen most intense ions from scans having either +2, +3, +4, or +5 charge states were selected for collision induced dissociation (CID) and tandem mass spectral (MS^2) detection (for review see Nunn and Timperman, 2007). Sample digests were analyzed using full scan (*m/z* 350-2000), followed by gas phase fractionation with repeat analyses over multiple narrow mass to charge ranges (e.g. *m/z* 350-444, 444-583, 583-825, 825-1600) (Yi et al., 2002; Nunn et al., 2006; Scherl et al. 2008).

Mass spectra were interpreted and searched using an in-house copy of SEQUEST on a Beowolf computer cluster with 800 dedicated processing cores and 22 terabytes of storage (Eng et al., 1994; Eng et al., 2008). All data searches were performed with no assumption of proteolytic enzyme (e.g. trypsin) specificity to allow for identification of the maximal number of protein degradation products. A fixed modification was set for 57 Da on cysteine (resulting from IAM alkylation step) and a variable modification of 16 Da on methionine via oxidation. Each tandem mass spectrum was then searched against a protein sequence database to correlate predicted peptide fragmentation patterns with observed sample ions. To objectively validate peptide and protein identifications, statistical evaluations using PeptideProphet and ProteinProphet were used to provide probability based scores (Keller et al., 2002; Nesvizhskii et al., 2003). Probability thresholds for positive identifications of proteins and peptides were strictly set at 90% confidence on ProteinProphet and PeptideProphet for SEQUEST search results. Mass spectra from all samples were searched against a database (referred to as "Thaps database") containing the proteomes of *Thalassiosira pseudonana* (marine diatom),

Prochlorococcus marinus (marine cyanobacterium), and *Pelagibacter ubique* (marine bacterium belonging to the SAR11 clade). The proteomes of *P. marinus* and *P. ubique* were included to account for potential input of bacterial proteins through the water column. The Thaps database was chosen after extensive comparison which revealed that larger databases, including the NCBI non-redundant database containing over 11.9 million protein sequences, did not enhance the number of protein identifications, added limited species diversity to identified proteins, and had 95% functional agreement between Thaps and larger database identified peptides from Bering Sea sediment (Moore et al., in review). This made the Thaps database suitable for accomplishing our goals of identifying and tracking algal proteins down the water column to sediments. False discovery rate was calculated to be 0.5% for Thaps database searches based on the identification rate of decoy peptide sequences.

A second separate database search was conducted to investigate correlations with a highly diverse assemblage of microbial peptides. This database included the proteome of *T. pseudonana*, plus the Global Ocean Survey (GOS) Combined Assembly Protein database (Rusch et al., 2007; Yooseph et al., 2007; Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis, CAMERA, downloaded on 29 September 2008). Although the GOS database has over 6,000,000 marine microbial proteins sequenced from genomic data, limited functional data is available. In this combined database inclusive of Thaps and GOS sequences, proteins from *T. pseudonana* account for approximately 0.1% of the total protein sequences, suggesting that randomly identified false spectra correlation to a *T. pseudonana* peptide was highly unlikely. Searches using the GOS/Thaps database were only completed on the post bloom surface

sediments where microbial products were expected to be most prevalent. The GOS/Thaps-searches were also limited because they were computationally intensive, consuming over 720 hours of search time (~1 month), even on the large computer cluster used. The intent of this search was to seek identifiable bacterial proteins in sediments which were expected to have high levels of microbial recycling.

Results

Protein distribution – Using a mass-spectrometry based approach, 207 proteins were identified in suspended particles from the chl-max of the Bering Sea (Table 4-2, Fig. 4-2A). Substantially fewer proteins were identified in suspended particles at 50 m (11 proteins) and 100 m (22 proteins). The number of identified proteins in the particle traps decreased from 136 at 40 m to 53 at 60 m, and 82 at 100 m. In post-bloom shelf sediment 52 proteins were identified, with slightly less than half the number of identifications in post-bloom basin and over-wintered shelf sediment. The majority of identified proteins at depth were also identified in the chlorophyll max, representing transport down the water column (Fig. 4-3). Sequence coverage, defined as the percentage of a specific protein sequence observed using tandem mass spectrometry, was highest on average in the chl-max and lowest in 50m POC.

The vast majority of peptides observed in all samples correlated to diatom proteins (Table 4-2). In the two samples with the highest number of protein identifications, chl-max particles and the 40 m particle trap, there were 200 and 129 diatom protein identifications respectively, plus seven bacterial identifications in each sample. Overall, the number of identified bacterial proteins accounts for 5% of the total identifications, with the remaining 95% identifications correlating to a diatom origin. Post-bloom shelf sediment mass spectra searched against the larger combined Thaps/GOS database yielded no unique bacterial protein identifications with confidence. The majority of identified bacterial protein's amino acid sequences overlapped with the amino acid sequences of *T. pseudonana* proteins as well.

Table 4-2. The number of total proteins identified in suspended particles (POC), particle traps, and sediments. Species distribution is based on database identifications originating from diatoms, (*T. pseudonana*), an autotrophic bacteria (*P. marinus*), or the pelagic bacteria (*P. ubique*).

Sample	Total Proteins Identified	*T. pseudonana*	*P. marinus*	*P. ubique*
Chl-max POC	207	200	3	4
50m POC	11	10	0	1
100m POC	22	19	0	3
40m Trap	136	129	5	2
60m Trap	53	52	1	0
100m Trap	82	79	3	0
PBS	52	49	1	2
PBB	24	21	1	2
OWS	23	22	1	0

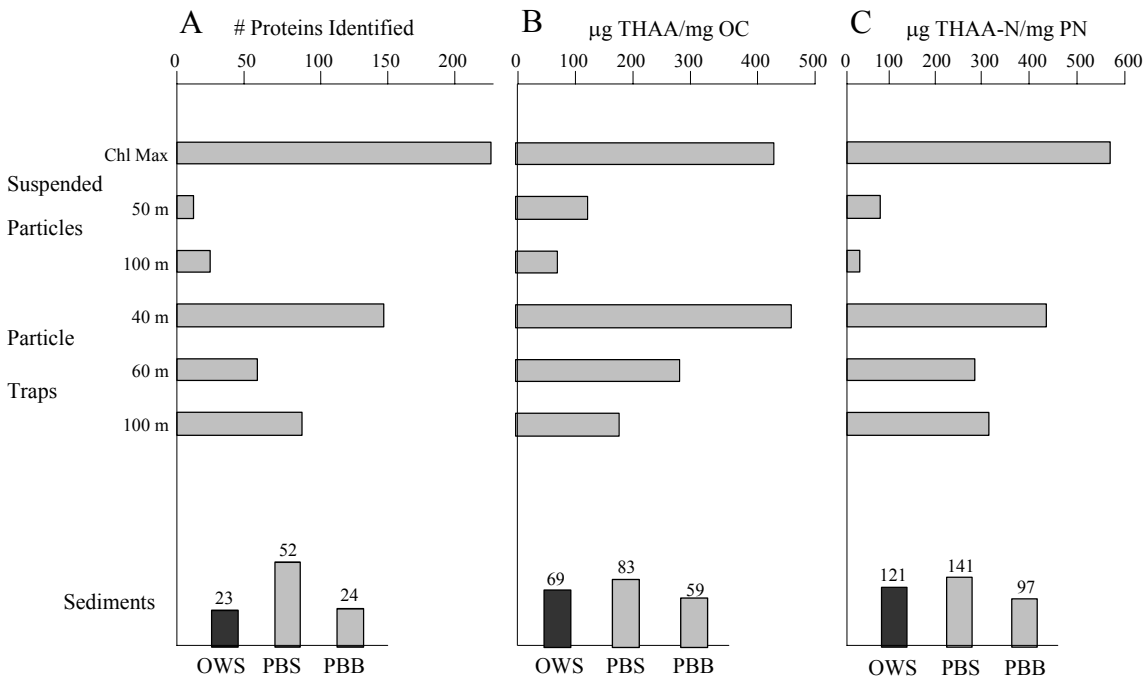

Figure 4-2. Profiles of (A) Number of proteins identified; (B) Ratio of total hydrolysable amino acids to organic carbon (THAA/OC) from each sample; (C) Ratio of total hydrolysable amino acid nitrogen to particulate nitrogen (THAA-N/PN) from each sample (Sediment samples graphed by residence time: OWS = Over-wintered Shelf; PBS = Post-bloom Shelf; PBB = Post-bloom Basin). Dark OWS bars represent older material collected before the spring phytoplankton bloom.

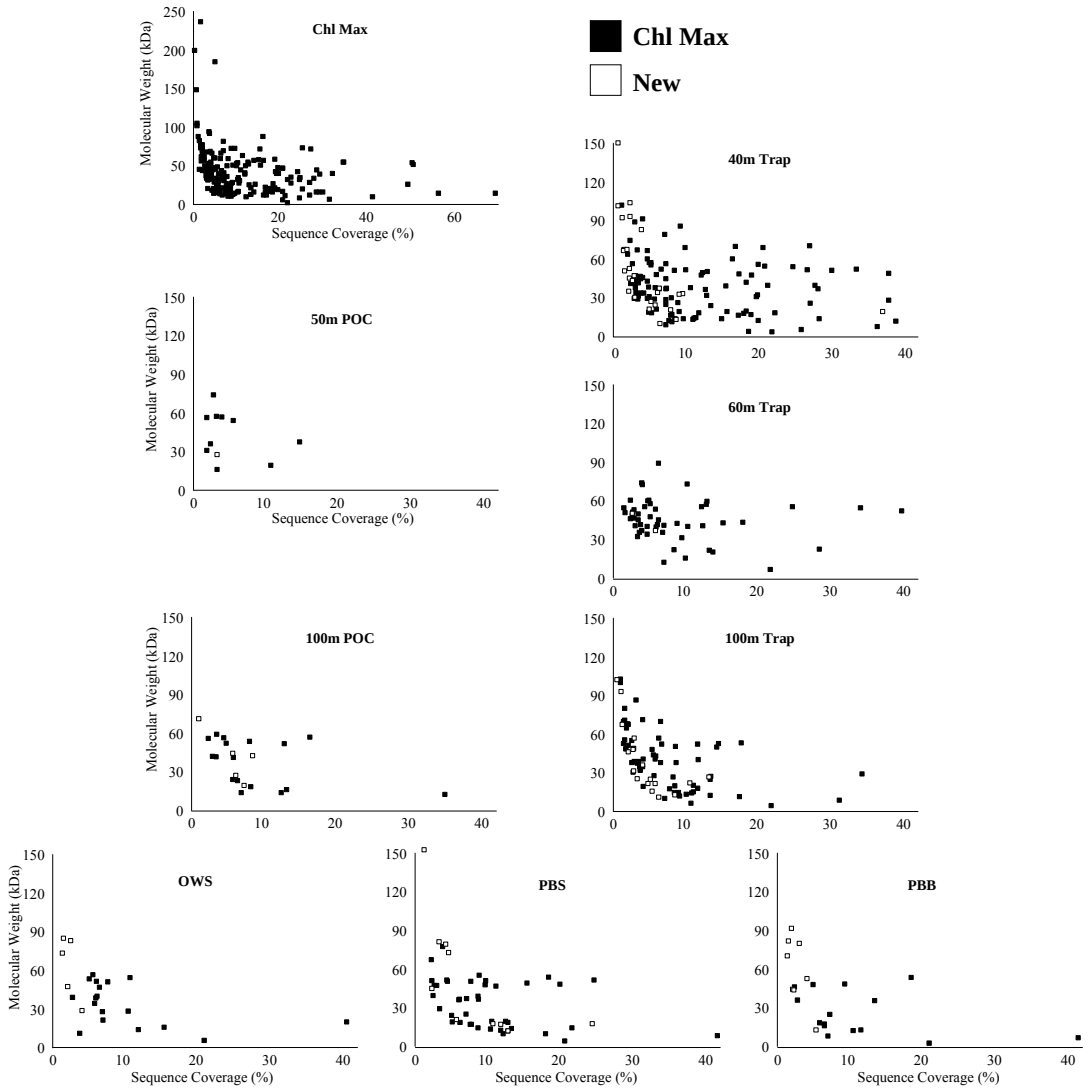

Figure 4-3. Protein molecular weight vs. sequence coverage plots for each sample. Solid data points represent proteins that were identified in chl-max, open data points represent proteins that were not identified in chl-max. One ~400 kDa protein (sequence coverage: 1%) in PBB plot has been excluded for ease of comparison.

As the number of identified proteins declined with increasing depth, the functional distribution of observed proteins changed. Protein categorization by gene ontology (Ashburner et al., 2000) revealed that metabolic proteins made up the largest functional group in chl-max particles at 63%, followed by structure/binding proteins at 18%, and transport proteins at 13% (Table 4-3, Appendix 4-1). A more even distribution of proteins among the metabolic, structure/binding, and transport groups was observed in 50 m and 100 m suspended particles. In all particle trap and sediment samples, metabolic proteins make up the largest group. Within the metabolic category, translation proteins were the largest subgroup in the chl-max at 17% followed by photosynthesis/carbon fixation proteins at 15% (Table 4-3). Deeper in the water column, photosynthesis/carbon fixation proteins accounted for 7% to 25% of identified proteins in suspended particles and particle traps; this fraction increased slightly in post-bloom shelf sediment to 26%, then considerably increased to 46% in post-bloom basin sediment and 52% in over-wintered shelf sediment. Conversely, the proportion of transport proteins decreased from water column suspended particles and traps to post-bloom basin and over-wintered shelf sediment. The percentage of proteins classified as structural or binding proteins did not change to the same degree as photosynthesis or transport proteins.

The comparison of protein abundance with more traditional measures found that THAAs decreased from the chl-max (520 μg/l) to 50 m suspended particles (14.2 μg/l), with a much smaller reduction from 50 m to 100 m (9.0 μg/l) (Table 4-1, Figure 4-2). Sediment trap THAAs decreased more steadily with depth than seen in suspended particles, not unlike identified proteins. Post-bloom shelf sediment had greater THAA concentrations than post-bloom basin sediment and over-wintered shelf sediment.

Table 4-3. The distribution of proteins observed in particles and sediments categorized by major cellular function as defined by Gene Ontology (Metabolic, Structure/Binding, or Transport). Total proteins observed are categorized as percentage of each by subgroup.

Chl-max POC IDs	207				
% Metabolic	131 (63.3%)	% Structure/Binding	37 (17.9%)	% Transport	27 (13.0%)
Translation	35 (16.9%)	Binding, DNA/RNA	13 (6.3%)	Ion Transport	19 (9.2%)
Photosynthesis	30 (14.5%)	Binding, Protein	8 (3.9%)	Transport, General	8 (3.9%)
Biosynthesis	18 (8.7%)	Protein Folding	7 (3.4%)		
Cellular Processing	29 (14.0%)	Binding, Molecule	6 (2.9%)		
Glycolysis	11 (5.3%)	Structural	3 (1.4%)		
Oxidation Reduction	8 (3.9%)				
50m POC IDs	11				
% Metabolic	3 (27.3%)	%Structure/Binding	2 (18.2%)	% Transport	5 (45.5%)
Photosynthesis	2 (18.2%)	Binding, Protein	1 (9.1%)	Ion Transport	4 (36.4%)
Translation	1 (9.1%)	Structural	1 (9.1%)	Transport, General	1 (9.1%)
100m POC IDs	22				
% Metabolic	8 (36.4%)	% Structure/Binding	7 (31.8%)	% Transport	7 (31.8%)
Photosynthesis	5 (22.7%)	Binding	5 (22.7%)	Ion Transport	5 (22.7%)
Cell Processing	3 (13.6%)	Structure	2 (9.1%)	Transport, General	2 (9.1%)
40m Trap IDs	136				
% Metabolic	82 (60.3%)	% Structure/Binding	25 (18.4%)	% Transport	24 (17.6%)
Photosynthesis	24 (17.6%)	Binding, DNA/RNA	9 (6.6%)	Ion Transport	19 (14.0%)
Translation	16 (11.8%)	Binding, Protein	7 (5.1%)	Transport, General	5 (3.7%)
Biosynthesis	12 (8.8%)	Protein Folding	3 (2.2%)		
Glycolysis	10 (7.4%)	Structure	3 (2.2%)		
Cellular Processing	16 (11.8%)	Binding, Molecule	3 (2.2%)		
Oxidation Reduction	4 (2.9%)				
60m Trap IDs	53				
% Metabolic	34 (64.2%)	% Structure/Binding	8 (15.1%)	% Transport	10 (18.9%)
Photosynthesis	13 (24.5%)	Binding, DNA/RNA	2 (3.8%)	Ion Transport	6 (11.3%)
Glycolysis	6 (11.3%)	Binding, Protein	3 (5.7%)	Transport, General	4 (7.5%)
Cellular Metabolism	5 (9.4%)	Protein Folding	2 (3.8%)		
Translation	4 (7.5%)	Structure	1 (1.9%)		
Biosynthesis	3 (5.7%)				
Cellular Processing	3 (5.7%)				
100m Trap IDs	82				
% Metabolic	39 (47.6%)	% Structure/Binding	12 (14.6%)	% Transport	18 (22.0%)
Translation	15 (18.3%)	Binding	7 (8.5%)	Ion Transport	16 (19.5%)
Photosynthesis	6 (7.3%)	Protein Folding	3 (3.7%)	Transport, General	2 (2.4%)
Glycolysis	6 (7.3%)	Structure	2 (2.4%)		
Cellular Metabolism	5 (6.1%)				
Cellular Processing	5 (6.1%)				
Biosynthesis	2 (2.4%)				
Post-bloom Shelf IDs	53				
% Metabolic	21 (39.6%)	% Structure/Binding	14 (26.4%)	% Transport	17 (32.1%)
Photosynthesis	14 (26.4%)	Binding, DNA/RNA	6 (11.3%)	Ion Transport	12 (22.6%)
Cellular Processing	4 (7.5%)	Structure	4 (7.5%)	Transport, General	5 (9.4%)
Translation	3 (5.7%)	Binding, Protein	4 (7.5%)		
Post-bloom Basin IDs	24				
% Metabolic	17 (70.8%)	% Structure/Binding	4 (16.7%)	% Transport	2 (8.3%)
Photosynthesis	11 (45.8%)	Binding, DNA/RNA	2 (8.3%)	Ion Transport	2 (8.3%)
Cellular Processing	4 (16.7%)	Structure	1 (4.2%)		
Translation	2 (8.3%)	Membrane	1 (4.2%)		
Over-wintered Shelf IDs	23				
% Metabolic	16 (69.6%)	% Structure/Binding	5 (21.7%)	% Transport	2 (8.7%)
Photosynthesis	12 (52.2%)	Binding, DNA/RNA	3 (13.0%)	Ion Transport	2 (8.7%)
Translation	2 (8.7%)	Binding, Protein	1 (4.3%)		
Cellular Processing	2 (8.7%)	Structure	1 (4.3%)		

Hydrophobic amino acids (Leu, Gly, Ala, Phe, Ile, Val) were the most represented amino acids in all samples, making up 56% of the average amino acid distribution in THAAs and 47% of the amino acid distribution of identified protein amino acid sequences (Fig. 4-4, Appendix 4-2).

Molecular weight distribution of surviving proteins – To compare the potential for selective loss of individual proteins based on molecular weight, identified proteins were grouped into five molecular weight ranges (<10, 10-25, 25-50, 50-100, >100 kDa) based on two categories. The first category was the predicted molecular weight of intact proteins as identified using the protein database search. The second was the molecular weight range in which each protein was observed following gel purification. Proteins identified in their anticipated gel molecular weight range can be categorized as "expected" since molecular weights were in agreement. In contrast, proteins observed by gel electrophoresis to be outside their predicted molecular weight range were categorized as "observed" in order to denote the disparity between gel mobility and expected molecular weight (Fig. 4-5). Of the proteins grouped by "expected" molecular weight, the vast majority of proteins identified in particles, traps, and sediments were in the intermediate size ranges (10-25, 25-50, 50-100 kDa) expected for many cellular proteins (Fig. 4-5A). The group with the most identifications comprises those proteins in particles and traps from 25-50 kDa. This changed as the proportion of 25-50 kDa proteins decreased in sediments compared to water particles and traps. The majority of "observed" proteins in suspended particles and particle traps were found to have greater than expected molecular weights ("observed larger") based on gel migratory behavior (Fig. 4-5B). The proportion of proteins with lower than expected molecular weights

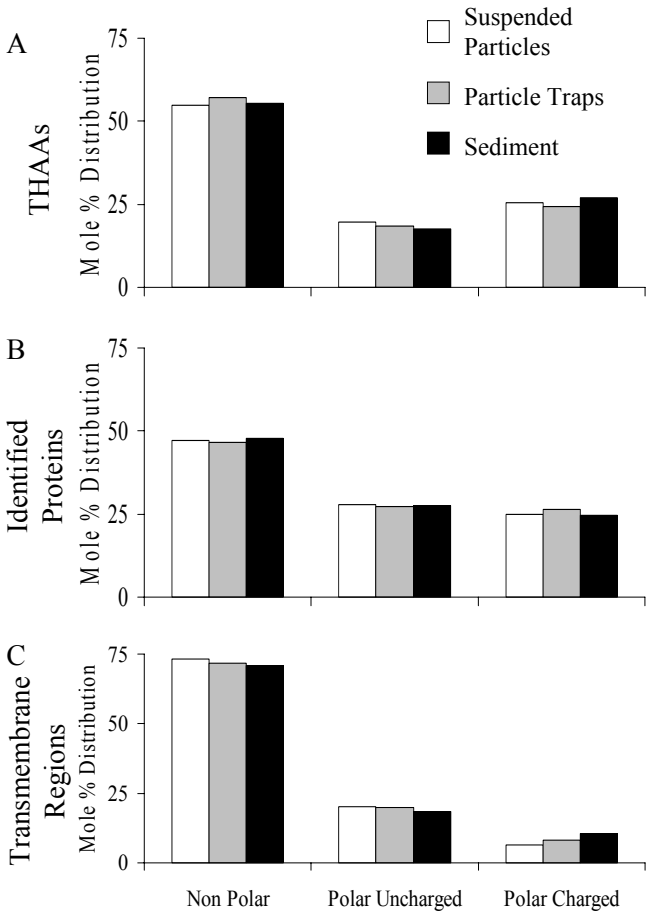

Figure 4-4. The distribution of non polar, polar uncharged, and polar charged amino acids among (A) total hydrolysable amino acids; (B) tabulated amino acids of identified proteins; (C) tabulated amino acids of TMHMM modeled transmembrane regions within identified proteins. Identified protein amino acids and THAAs show very similar distribution, while transmembrane regions have greater proportion of non polar amino acids.

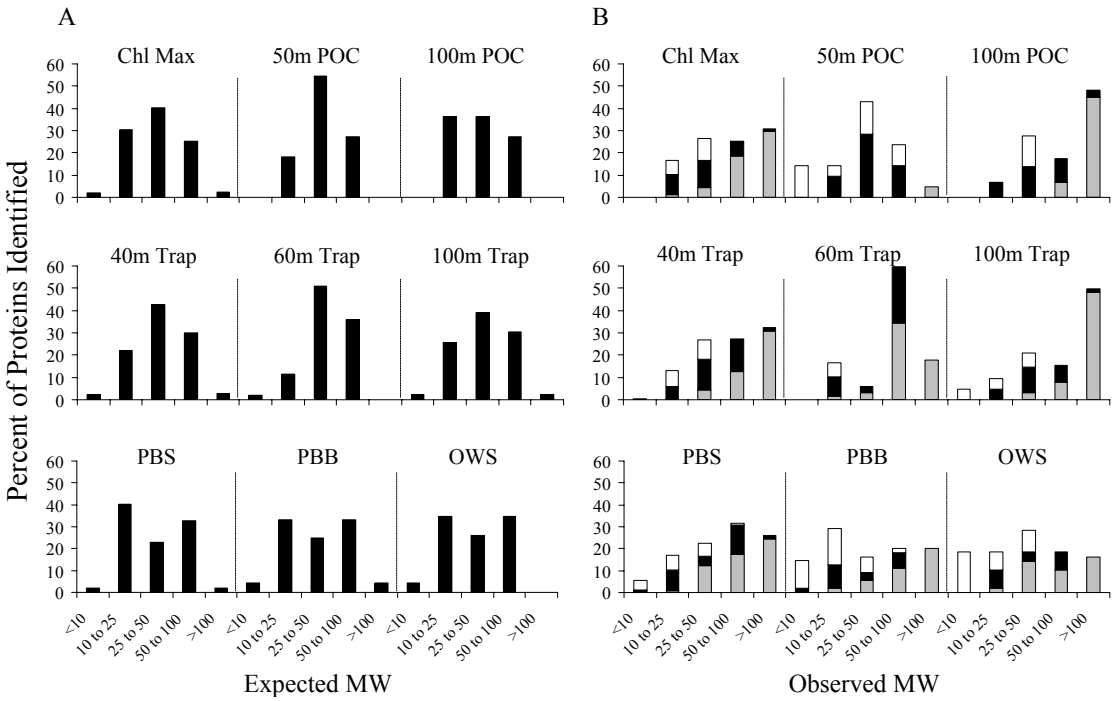

Figure 4-5. The percentage of identified proteins in each sample within molecular weight groups verses identified proteins with altered properties. A) The expected distribution of identified proteins in samples; (B) The distribution of proteins identified in each gel section: black bars represent the percentage of proteins identified in the expected gel section molecular weight range, grey bars represent "observed large" proteins, white bars represent "observed small" proteins.

("observed smaller") increased in sediments, especially post-bloom basin and over-wintered shelf sediments.

Discussion

The identification of unique peptide sequences allow the demonstration that proteins derived from primary production are able to survive the degradation processes during initial transit to the sediment water interface. Over 70% of all proteins identified in suspended particles at depth, and over 80% of proteins identified in particle traps were also detected in the diatom dominated chlorophyll maximum. The importance of primary production proteins extended to both shelf sediment samples, representing >80% of identified proteins, and basin sediments where 67% of identified proteins were identical to those observed in the chl-max. Diatoms may contribute organic matter to sediments because they are encapsulated in high-density silica frustules that can rapidly transport organic matter to depth (Dunne et al., 2005; Ragueneau, et al., 2006; Miki et al., 2009). Multiple fucoxanthin chlorophyll a/c binding proteins (FCPs), important light harvesting complex proteins in diatoms and other marine algae (Grossman et al., 1995; Lang and Kroth, 2001; Nunn et al., 2009), were observed in all sediment samples. The presence of these proteins is not completely unexpected; FCPs are central in the light harvesting complex, representing the most abundant protein class discovered in mid-exponential growth *T. pseudonana* (Nunn et al., 2009) and later observed to remain after extensive microbial attack in a controlled month-long degradation experiment (Nunn et al, 2010) . Since diatoms dominate spring bloom production in the Bering Sea, and their density and aggregation result in sinking after bloom termination (Smetacek et al., 1985), it is very likely that diatoms are the source of algal-derived light harvesting proteins to Bering Sea sediments.

The discovery of diatom and algal specific proteins in ocean basin sediments reveals that, despite the intensive recycling processes active in oceanic waters, highly organized macromolecules such as proteins can be transported intact or partially degraded from their biosynthetic origins in surface waters to sediments. While fewer proteins were identified in this study than Dong et al. (2010), greater statistical rigor was applied by way of PeptideProphet and ProteinProphet to identify peptides and proteins in the system with greater confidence. In addition, the diatom/algal proteins identified were not only observed over the shallow productive continental shelf, but also in the deeper ocean surface sediments, suggesting that seasonal primary production in the Bering Sea is also an important contributor to deeper environments. Given that diatom sinking rates can range from 40 m per day to over 100 m per day (Smetacek, 1985 and references therein), the sinking time for bloom material to the shelf sediment-water interface (101-136 m) would be on the order of days, while sinking time over the basin (3490 m) would be on the order of weeks. Recent laboratory studies by Nunn et al., (2010) found that a subset of diatom proteins can be retained over a 23 day degradation period, which could encompass the potential sinking time of bloom material to sediments and initial sedimentary recycling. Many more proteins were identified in sinking sediment trap material than in suspended particle samples (Table 4-2), reinforcing previous observations that sinking material has a distinct composition compared to suspended particles (Sheridan et al., 2002; Abramson et al., 2010). To our knowledge, this is the first demonstration that intracellular, cytosolic and soluble proteins can be transported to depth from sources in the upper water column. All these factors suggest that large dense diatom cell size and composition is central in transport, burial, and eventual preservation

of sedimentary proteins from a range cellular compartments in the Bering Sea and other diatom dominated systems.

It has long been observed that as marine organic matter becomes more degraded, there is a decrease in the ratios of total hydrolysable amino acids to organic carbon (THAA/OC), and THAA-nitrogen to particulate nitrogen (THAA-N/PN) (e.g. Cowie and Hedges, 1994 and references therein). Here, these two ratios were used to normalize amino acid amounts across multiple sample types, and compare amino acid concentrations with the number of identifiable proteins in each sample. Plotting the number of identified proteins against the ratios of THAA/OC and THAA-N/PN, strong correlations are observed for both carbon and nitrogen (Fig. 4-6A, B). The correlation between protein IDs and THAA-N/PN is particularly strong, displaying the importance of protein to marine organic nitrogen (Brown, 1991; Lourenco et al., 1998). These relationships also show that over a wide range of concentrations, THAAs present in particles and even surface sediments likely include a portion of intact proteins, with the number of identified proteins reflecting the extent to which proteins have been degraded during diagenesis. In the case of this study where diatoms are the primary source material, the relationship likely exhibits the freshness of produced organic matter and retention of diatom proteins.

Arranging samples in order of decreasing THAA-N/PN values, sample clusters are formed: chl-max particles > all particle traps > deeper suspended particles and surface sediment, which likely represents increasing degradation status (Fig. 4-6B). This order also generally reflects the amount of identifiable proteins in each sample and average sequence coverage of identified proteins in most samples. Fewer identifiable proteins

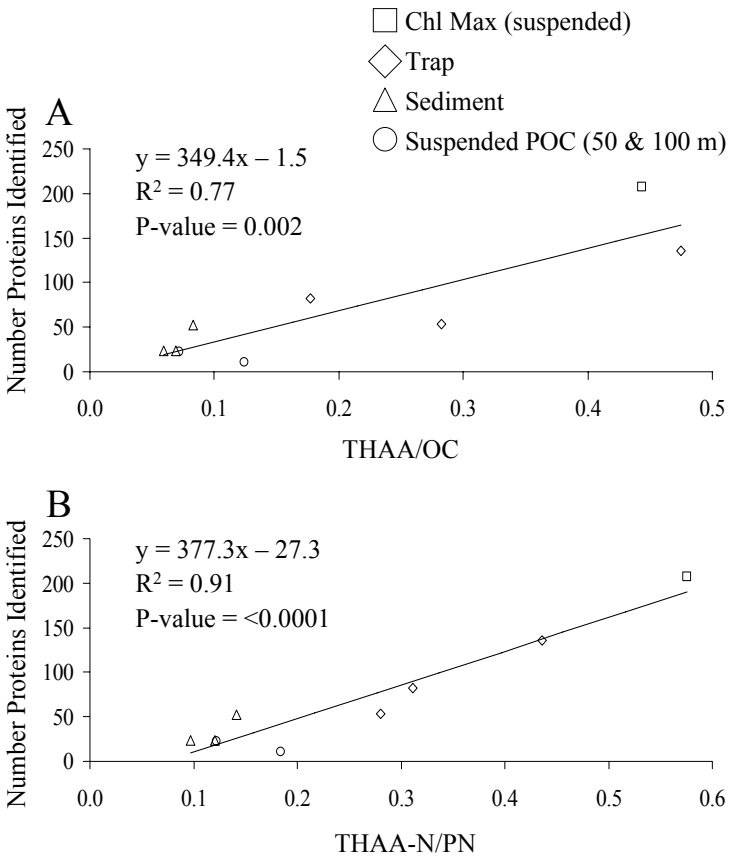

Figure 4-6. (A) The number of identified proteins plotted against THAA/OC; (B) Identified proteins plotted against THAA-N/PN with linear regressions displayed for each.

and similar THAA-N/PN proxies from suspended particles suggest a similar "degraded" status for proteins compared to those found in the sediments. This may reflect the length of time since organic matter production (Hartnett et al., 1998) as well as the potential importance of sorption to sediment for protein preservation (Collins, 1995; Mayer, 1999). The enhanced sedimentation rates of large aggregates could control the fraction of proteins present in surface sediments.

In addition to the absolute number of identified proteins and THAAs, the diversity of different protein groups based on cellular function declined in deeper waters and with residence time in surface sediments. Metabolic proteins made up the largest group in all samples except for the 50 m suspended particle sample (Table 4-3). Within the metabolic protein category, photosynthetic and carbon fixation proteins represent the largest percentage of proteins remaining in sedimentary material. Recent work by Nunn et al. (2010) highlighted factors that might influence the preservation of such proteins over short time scales. Several characteristic traits were proposed to encourage protein stability and/or longevity including organelle compartmentalization, transmembrane-spanning domains, initial cellular abundance, glycan modifications, and aggregation. Several of these mechanisms can be considered in the context of Bering Sea protein preservation.

Compartmentalization and preservation potential – Proteins are not uniformly distributed in cells, but typically associated with various cellular compartments which might influence preservation and help explain the observed distribution shifts of protein cellular functions through water column loss. Gel electrophoresis has been shown to enhance liberation of membrane bound proteins (Coughenour et al., 2004) allowing

improved identification of organelle proteins in addition to molecular weight separation. To examine the possibility of organelle preservation, protein compartments and subcellular locations were assigned using the TargetP 1.1 Server (Emanuelsson et al., 2007) and uniprot.org (Jain et al., 2009) respectively based on the *T. pseudonana* proteome (Fig 4-7).

The proportion of membrane compartmentalized proteins discovered in sediments (i.e. chloroplast and mitochondrial proteins) increased from post-bloom shelf sediment to both post-bloom basin and over-wintered shelf sediments, while the proportion of secretory proteins decreased. Similarly, the percentage of identified proteins with transmembrane regions, as modeled by TMHMM Server v. 2.0 (Krogh et al., 2001) increased in deeper particles and sediments (Table 4-4). These observations suggest that soluble secretory proteins, with no added organelle membrane protection, are more susceptible to microbial recycling during sinking to basin sediments and over longer sediment residence time in over-wintered shelf sediment compared to shorter sinking and residence time experienced by post-bloom shelf sediment material. Similar patterns were recently seen in laboratory incubations (Nunn et al. 2010) with the preferential preservation of organelle enclosed proteins. This includes the important light harvesting FCP proteins which are enclosed in 4 membrane layers of the thylakoid and chloroplast (Westermann and Rhiel, 2005). Previous work has shown proteins enclosed in crude membrane extracts are degraded more slowly compared to soluble proteins (Nagata et al., 1998), and protein sequences derived from conserved membrane/envelope proteins were previously identified from Gulf of Mexico DOM (Powell et al. 2005). These lines of

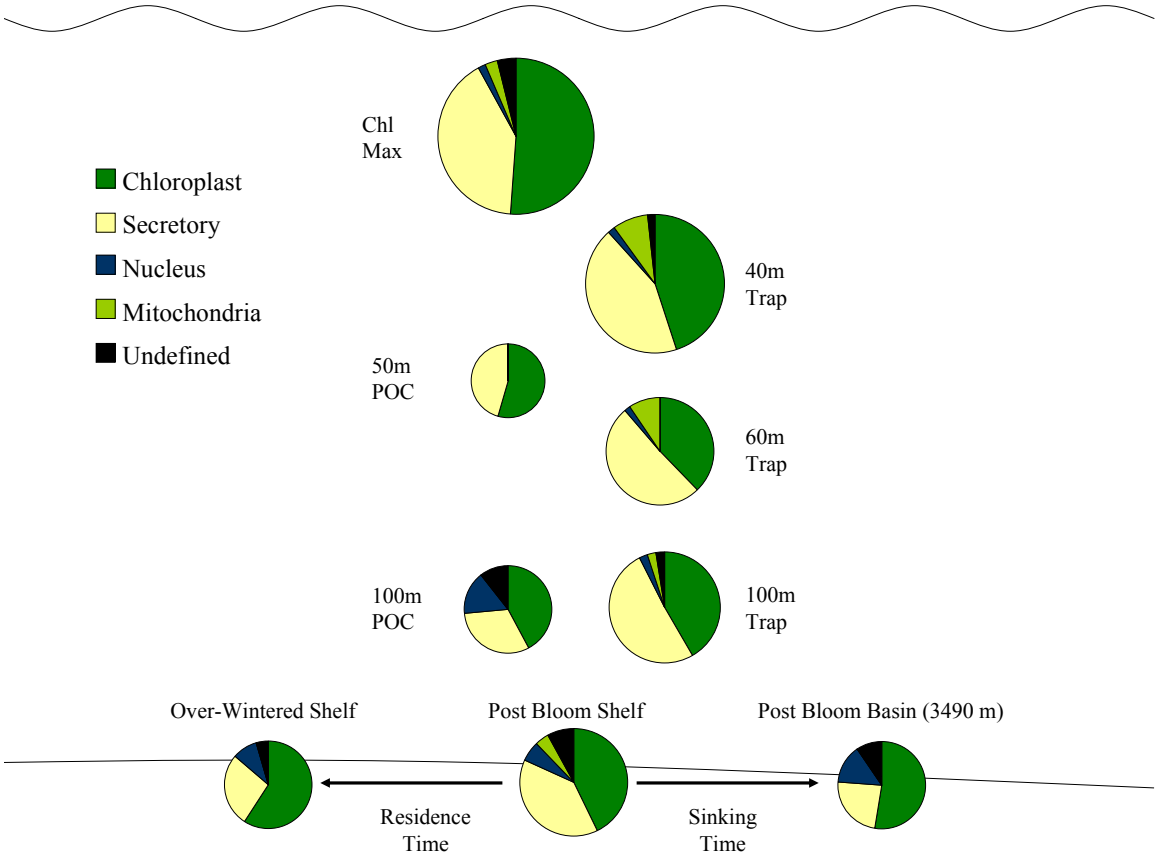

Figure 4-7. The relative distribution of proteins identified from the cellular compartments of diatoms. Each section shows the percent contribution of proteins originating from a specific compartment. Cellular compartments include: chloroplast, secretory, nucleus, mitochondria, and a group of proteins that are from undefined or unknown compartments. The size of each circle is scaled to the number of proteins identified in each sample.

Table 4-4. The average isoelectric point (pI), number and relative abundance of high abundance diatom proteins[1], number and percentage of transmembrane proteins, percent transmembrane amino acids from each sample, and number and percentage of transmembrane proteins located in the chloroplast. Transmembrane regions predicted by TMHMM server v. 2.0.

Sample	pI	High Abundance Proteins[1]	Membrane Proteins	Percent Transmembrane AAs	Transmembrane Proteins in Chloroplast
Chl Max	6.5	59 (29%)	25 (12%)	2	17 (68%)
50m POC	5.6	8 (80%)	2 (18%)	16	2 (100%)
100m POC	6.8	13 (59%)	6 (27%)	8	4 (67%)
40m Trap	6.2	48 (37%)	20 (15%)	3	15 (75%)
60m Trap	5.5	30 (58%)	5 (9%)	2	5 (100%)
100m Trap	6.6	32 (41%)	16 (20%)	5	13 (81%)
PBS	6.9	23 (47%)	15 (28%)	9	12 (80%)
PBB	6.8	9 (43%)	9 (38%)	9	8 (89%)
OWS	6.8	12 (55%)	10 (44%)	14	9 (90%)

[1] high abundance proteins were defined as the top 1% of proteins based on proteome analysis (Nunn et al., 2009)

evidence point to physical protection through membrane associations as a factor in the survival of marine proteins.

Implications of identified proteins on THAA distribution – Although early reports found that the relative distribution of particulate amino acids show only minor changes with ocean depth (Wakeham et al., 1984; Müller et al., 1986), or as algae are subject to degradation (Nguyen and Harvey, 1997), amino acid distributions have been used as a metric for degradation state (Dauwe and Middelburg, 1998 and references therein). Given that the amino acid distribution of most proteins is very similar (Brooks et al., 2002), such changes are small. We can compare total THAAs among all sample types with amino acids tabulated from the sequences of identified proteins and those amino acids associated only with transmembrane regions of proteins (Fig. 4-4A-C). Grouped for ease of comparison into major functional groups, the distribution of amino acids tabulated from identified proteins shows remarkable similarity to THAAs among suspended, sinking, and sedimentary material. The proportion of hydrophobic amino acids was highest for protein transmembrane domains among all sample types (71-73% non-polar), and may reflect hydrophobic interactions previously observed as stabilizing forces for proteins in particulate material and organic rich sediments (Nguyen and Harvey, 2001) and kerogens (Nguyen and Harvey, 2003).

Protein abundance and identification – The array of possible marine proteins is vast, as reflected by the breadth of the GOS protein database (Yooseph et al., 2007). Identification of proteins by data-dependent ion selection of peptides during HPLC-MS/MS is inherently biased to more abundant peptides in complex matrices and thus proteins in low concentrations may remain below detection limits. This suggests that the

most abundant proteins present in diatom cells are more likely to be available for identification after long term microbial attack. To test this, proteins ranked in the top ~1% of abundance based on the proteome analysis of Nunn et al., (2009) were classified as high abundance proteins, while the remaining proteins in the *T. pseudonana* proteome were grouped as low abundance proteins. The proportion of high abundance proteins increases among identified proteins with depth from bloom material to sinking trap material to sediments (Table 4-4). The percentage of high abundance proteins was also greater in over-wintered shelf sediment than either post-bloom or deeper shelf sediment samples. This suggests that initial abundance of individual proteins in living cells influences the potential for their detection after substantial losses via degradation. Once a protein reaches the sediment, preservation mechanisms such as aggregation (Nguyen and Harvey, 2001; 2003), particle (Nagata and Kirchman, 1996), or mineral sorption (Mayer, 1994; 1999; Hedges and Keil, 1999), are likely to extend its longevity.

Protein molecular weight – An important caveat to the identification of proteins is that absolute molecular weight is not measured, and thus proteins which are retained in the system might represent non-native forms. Indeed, covalent modifications have been previously proposed as one mechanism for protein preservation in older sediments (Cronin and Morris, 1981; Benner et al, 1992; Nguyen and Harvey, 2003). Given the molecular weight distribution of low, intermediate, and high molecular weight proteins in the *T. pseudonana* proteome (Armbrust et al., 2004; Nunn et al., 2009), the observed electrophoretic behavior among a subset of identified proteins suggests that changes to the original sequence (or charge) are common and can lead to higher than expected molecular weights based on mobility within a gel (Fig. 4-5B). While aggregation has

been proposed previously as one mechanism (Nguyen and Harvey, 2001; 2003), the denaturing conditions used here suggest that covalent modifications or charge alterations are also operative. The fact that there were also a significant number of proteins which appear to have lower molecular weights than expected suggests that partial hydrolysis is active as well (i.e. Pantoja and Lee, 1999).

Bacterial proteins in sediment – In the Bering Sea as in other ocean environments, microbial processes act as the primary catalyst for organic matter recycling. Although estimated bacterial biomass present in sediments is far lower than that derived from primary production, proteins associated with active microbial populations undoubtedly are present. Mass spectra collected from Bering Sea shelf post-bloom sediments searched against the *T. pseudonana*/GOS Combined Assembly Protein database yielded only two bacterial protein identifications at the 90% confidence level: 1) a vacuolar proton inorganic pyrophosphatase and 2) translation elongation factor 1 alpha. The spectra that correlated to peptides from the GOS database from these two proteins also correlated with known *T. pseudonana* peptides, demonstrating that in many cases different species have homologous peptide sequences or that sequences are not well constrained.

Challenges with identifying bacterial proteins noted in the soil literature include high diversity of known and unknown bacterial proteins resulting in incomplete bacterial protein databases and individual proteins being below detection limits (Graves and Haystead, 2002, Quince et al., 2008; Bastida et al., 2009). Furthermore, bacterial proteins may be masked by co-extracted sedimentary material (Criquet et al., 2002) confounding the issue of detection limits missing highly diverse low abundance proteins.

Nevertheless, proteomic studies on soil organic matter have revealed functional information on microbial communities and particle bound material (Schulze et al, 2005; Benndorf et al., 2007). Bacteria appear principally as catalysts in this first examination of protein distributions during organic matter recycling in marine systems, with phytoplankton derived material as the confirmed source of proteins in Bering Sea sediments. Simple estimations based on typical bacterial cell abundance in marine sediments (~10^9 cells g^{-1}; Griffiths, et al., 1978; Deming and Colwell, 1982; Harvey, et al., 1984; Luna, et al., 2002; Kopke, et al., 2005) and average protein content per cell (24 fg $cell^{-1}$, Zubkov et al., 1999) indicate that bacteria contribute approximately 3% compared to THAAs used as a proxy for total sedimentary protein. Estimations of protein content in sediments based on particle flux measurements (Takahashi et al., 2002), carbon content, and THAA content suggested that annual algal inputs accounted for approximately 23% of sedimentary protein material. Considering that algal protein can survive seasonal timescales and accumulate in sediments, it appears that bacteria contribute a small fraction to the total protein pool examined here.

Conclusions

This is the first study to apply an MS-based proteomics approach to follow the environmental fate of phytoplankton specific proteins present during bloom, transport, and initial incorporation into sediments. The survival of algal derived proteins appears selective, with compartmentalized and cell membrane proteins demonstrating greater longevity after genesis and short-term recycling in the water column. The correlation between identifiable proteins and THAA-N/PN suggests that proteins are not all rapidly hydrolyzed but may represent a predictable fraction of organic nitrogen present in organic matter. While database and detection limit challenges hinder the identification of bacterial proteins, phytoplankton proteins appear to be important contributors to Bering Sea sedimentary THAAs.

Chapter 5: Protein preservation over 11 and 53 day Bering Sea algal decay incubations: pre and post death proteome changes

Abstract

Protein from phytoplankton blooms represents a large fraction of new organic nitrogen and carbon that undergoes recycling and transport to marine sediments following cell death. However, little is known about the recycling timelines of identifiable protein in situ or the potential modifications to proteins that occur during bloom termination. To address this, phytoplankton bloom material was collected in the Bering Sea during the spring of 2009 and 2010, and incubated under darkness in separate shipboard degradation experiments spanning 11 and 53 days respectively. The distribution of proteins was monitored over the course of the incubations using shotgun proteomics, along with total hydrolysable amino acids (THAAs), total protein, particulate organic carbon (POC), particulate nitrogen (PN), and bacterial cell abundance. After the initial loss of identifiable proteins, total protein, and THAAs, recycling proceeded at a slower rate after 5 days in both incubations. Algal proteins were identified after 53 days of degradation confirming previous observations that protein can survive time frames that allow transport to marine sediments. The proportion of proteins involved in photosynthesis increased as the number of identifiable proteins declined through both incubations. In addition, modified peptides were identified from samples that were first digested with PNGase F to remove glycan modifications.

Introduction

Diatoms typically dominate phytoplankton biomass in the Bering Sea (Banahan and Goering, 1986; Springer et al., 1996; Sukhanova et al., 1999; 2009), one of the world's most productive ecosystems (Sambrotto et al., 1986; McRoy et al., 1987; Walsh et al., 1989). High export in this region (Chen et al., 2003), reduced impact by zooplankton grazing during high vertical flux periods after bloom termination, and the shallow average shelf depth combine to allow a large fraction of primary production to be transported to sediments with resulting high benthic biomass (Highsmith and Coyle, 1990; Lovvorn et al., 2005). Phytoplankton proteins have been tracked down the water column in sinking particles to 136m shelf and 3490m basin sediments in the Bering Sea, confirming the transfer of proteins from primary producers to the ocean floor (Moore et al., in reivew). Diatom aggregates have been estimated to sink at rates of 100m day^{-1} or greater (Smetacek, 1985), and this raises the question of how long phytoplankton protein material might remain during transport to sediments and before water column recycling takes place.

In a laboratory based system, diatom proteins have been observed over a 23 day microbial degradation experiment at 19^{o}C (Nunn et al., 2010). The greatest loss in the number of identifiable proteins took place during five day darkness period before bacteria were introduced for the 23 day degradation period. This suggests that diatoms restructured and recycled their proteome in order to acclimate to low light levels and disable replication pathways. For this reason, monitoring the early stage of diatom introduction into darkness may be just as important in understanding protein cycling as the bacterial degradation process. Various proteomic studies have been conducted on

plant programmed cell death (Chen et al., 2009; Choi and Hwang, 2011; Chivasa et al., 2011) and algal cell stress (Jamers et al., 2009; Contreras et al., 2010; Cui et al., 2010) identifying potential indicators and regulators of proteomic alteration. The shotgun proteomic analysis of Bering Sea bloom material during degradation allows the observation of proteome changes, selective preservation, and longevity of bloom material in environmentally relevant conditions from the water column. In a highly productive seasonal phytoplankton bloom population markers for cell death and subsequently bloom termination, and preserved protein structures and motifs throughout recycling will be useful in predicting organic nitrogen and carbon export by algal colonies and cell aggregates.

Methods

Incubation procedure – Bering Sea water was collected on the outer shelf during the spring bloom of 2009 and 2010. Twenty liter carboys were filled with CTD water from the chlorophyll maximum. Phytoplankton material was concentrated approximately 2X using 10 μm mesh and added to the 20 l carboys to increase amount of bloom material in the incubations. Carboys were placed in -1^{o}C cold rooms for 11 days (short incubation, 2009) and 53 days (long incubation, 2010). The cold rooms were illuminated with dim red light during sampling to minimize light exposure during incubation, and the carboys were covered with plastic garbage bags throughout the incubations. One liter water samples were collected periodically from each carboy and filtered onto 25 mm glass fiber filters (GF/F) and 37 mm polycarbonate 0.2 μm filters for analysis (Table 5-1). Five ml samples were collected at each time point and filtered through 0.2 μm, DAPI stained, and fixed onto microscope slides for bacterial counts (Kapuscinski, 1995). All incubation particles and bacterial slides were stored at -70^{o}C until analysis.

Amino acid and bulk analysis – Total hydrolysable amino acids (THAAs) were identified and quantified by gas chromatography/mass spectrometry (GC/MS) using the EZFaast method (Phenomenex ®) which uses derivatization of amino acids with propyl chloroformate and propanol for detection (Waldhier et al., 2010). Briefly, suspended particles collected on GF/Fs were hydrolyzed for 4 hours at 110 ^{o}C (Cheng et al., 1975; Cowie and Hedges, 1992) with 6 M analytical-grade HCl and L-γ-Methylleucine as the recovery standard. Following hydrolysis and derivatization, amino acids were quantified using an Agilent 6890 capillary GC with samples injected at 250 ^{o}C and separated on a DB-5MS (0.25 mm ID, 30 m) column with H_2 as the carrier gas. The oven was ramped

Table 5-1. Carbon to nitrogen ratio, total hydrolysable amino acids (THAAs), total protein (Bradford assay), bacterial cell abundance, and total protein identifications (using trypsin, EndoGluC, and PNGase + trypsin) from the 11 day and 53 day incubations.

11 Day Incubation

Days	C:N	THAAs (mg/l)	Total Protein (mg/l)	Bacteria (cells/l)	Total IDs
0	3.99	1013.00	1067.53	6.03E+07	127
0.5	3.63	448.52	618.51	-	48
1	3.77	612.73	545.53	5.60E+07	64
3	3.54	571.18	680.93	1.03E+08	41
5	3.58	539.73	568.45	1.62E+08	51
7	3.56	454.31	566.94	2.10E+08	37
9	3.59	424.91	386.81	2.73E+08	42
11	3.55	400.61	404.50	1.14E+08	39

53 Day Incubation

Days	C:N	Amino Acids (mg/l)	Total Protein (mg/l)	Bacteria (cells/l)	Total IDs
0	4.37	805.63	988.57	7.56E+07	95
5	4.57	439.98	453.59	1.63E+08	32
12	4.47	386.44	177.42	8.90E+07	31
22	4.62	337.93	233.63	7.94E+07	15
35	4.93	244.67	248.82	7.79E+07	13
47	4.96	200.72	185.55	8.04E+07	8
53	4.83	178.81	295.31	7.74E+07	13

from an initial temperature of 110 ^{o}C to 280 ^{o}C at 10 ^{o}C per minute followed by a 5 minute hold. Amino acid identification was accomplished by an Agilent 5973N mass spectrometer run under the same conditions with helium as the carrier gas and spectra acquisition over the 50-600 Da range. Bovine serum albumin (BSA) was analyzed in parallel to correct for responses among individual amino acids and calculation of molar ratios. Amino acids were normalized to percent carbon and nitrogen using bulk samples analyzed by standard combustion methods. Total protein content was estimated by the Bradford Assay. Stained bacterial cells were counted on an Olympus BH2-RFCA fluorescent microscope following shipboard staining and mounting.

Digestion of incubation samples – Incubation particles on Polycarbonate and GF/F filters were pulse sonicated in 6 M urea with a Branson 250 sonication probe at 20 kHz for 30 seconds on ice. The extracts were then frozen at -70 ^{o}C, thawed, and sonicated again for 30 seconds on ice. There were five sonications and four freeze/thaws in all. Filter extracts of each incubation time point were then digested in three replicate groups: 1) standard tryptic digestion with reduction and alkylation (Shevchenko et al., 1996); 2) digestion with Endoproteinase GluC (Endo GluC), which cleaves peptide bonds C-terminal to glutamic acid (Drapeau et al., 1972) and to a lesser extent aspartic acid (Birktoft and Breddam, 1994) to increase the number of proteins identified; 3) prior to tryptic digestion, extract was incubated with Peptide N-Glycosidase F (PNGase F), which hydrolyzes nearly all types of N-glycan chains from glycoproteins and glycopeptides (Maley et al., 1989) in order to detect peptides with potential glycan modifications. Digests were dried by vacuum concentration (Speedvac) to a volume that gave a final

protein concentration of approximately 1 μg/10 μl based on measured protein concentrations of filter extracts.

Mass spectrometry and database searching - Protein identification of sample digests was performed via shotgun proteomic tandem mass spectral (MS^2) detection (Aebersold and Mann, 2003). Digests were analyzed using full scan (*m/z* 350-2000), followed by gas phase fractionation with repeat analyses over multiple narrow, but overlapping mass to charge ranges (Yi et al., 2002; Nunn et al., 2006; Scherl et al. 2008). Mass spectra were evaluated and database searched with an in-house copy of SEQUEST (Eng et al., 1994; Eng et al., 2008). All searches were performed with no assumption of proteolytic enzyme cleavage (e.g. trypsin, Endo GluC) to allow for identification of protein degradation products due to microbial recycling. A fixed modification was set for 57 Da on cysteine and a variable modification of 16 Da on methionine resulting from alkylation and reduction steps respectively. A variable 1 Da modification was set for asparagine on PNGase F + trypsin digested samples to account for the conversion of asparagine to aspartic acid after cleavage of glycan chains with the use of PNGase F (Plummer et al., 1984; Chu, 1986), which takes place specifically at the consensus sequence Asn-Xxx-Ser/Thr where Xxx can be any amino acid excluding proline (Bause and Hettkamp, 1979; Kornfeld, 1985).

Each tandem mass spectrum was searched against a protein sequence database to correlate predicted peptide fragmentation patterns with observed sample ions. Probabilistic scoring of protein identifications was given by PeptideProphet and ProteinProphet (Keller et al., 2002; Nesvishskii et al., 2003). Thresholds were set at 90% confidence on PeptideProphet and ProteinProphet parameters for positive protein

identifications from SEQUEST search results. Mass spectra from all samples were searched against a database containing the proteomes of *Thalassiosira pseudonana* (marine diatom), *Prochlorococcus marinus* (marine cyanobacterium), and *Candidatus Pelagibacter ubique* (marine bacterium belonging to the SAR11 clade) which was termed the Thaps database. These proteomes were selected to observe protein degradation in a diatom dominated system with potential input of bacterial proteins, as in the diatom degradation experiment by Nunn et al., 2010. Database comparisons in previous studies showed functional agreement for over 95% of identified peptides between the *T. pseudonana-P. marinus-C. P. ubique* database vs. the larger Global Ocean Sequence (GOS) and NCBI non redundant databases containing over 6 million and 11 million protein sequences respectively (Moore et al., in review).

Results

Algal proteins, identified as originating from *Thalassiosira pseudonana,* were detected throughout both incubations, including after 53 days (Table 5-1, Figure 5-1A). After initial rapid decrease in identifiable proteins for both the 11 day and 53 day incubations, the rate of protein identification (ID) loss slowed after 5 days in both incubations. The number of protein IDs remained relatively consistent after 22 days in the 53 day incubation. Similar trends were observed in both incubations for THAAs (Figure 5-1B) and total protein (Figure 5-1C). The THAA distribution was fairly consistent among the 11 and 53 day incubations (Appendix 5-1), with a sharp drop in Alanine from day 0 to 0.5 in the 11 day incubation, and a spike in Glutamic Acid/Glutamine at day 22 of the 53 day incubation. Bacterial cell counts peaked during the first 9 days of both incubations about the same period that initial loss of protein IDs, THAAs, and total protein were most rapid (Figure 5-1D). Despite the increase in bacterial cell numbers, few prokaryotic proteins were identified over the course of the two incubations.

Chloroplast proteins and secretory proteins were the two largest cellular compartment groups of identified proteins in both incubations, with much smaller contributions of identified proteins from the mitochondria, nucleus, ribosome, and unknown compartments (Figure 5-2). As fewer proteins were identified in later time points of both incubations the sequence coverage of identified proteins generally decreased as well. Chloroplast proteins were the most persistent among all other cellular compartments increasing from 44% and 48% of total identifications at day 0 to 74% and 67% in the final time points of the 11 and 53 day incubations respectively. Conversely,

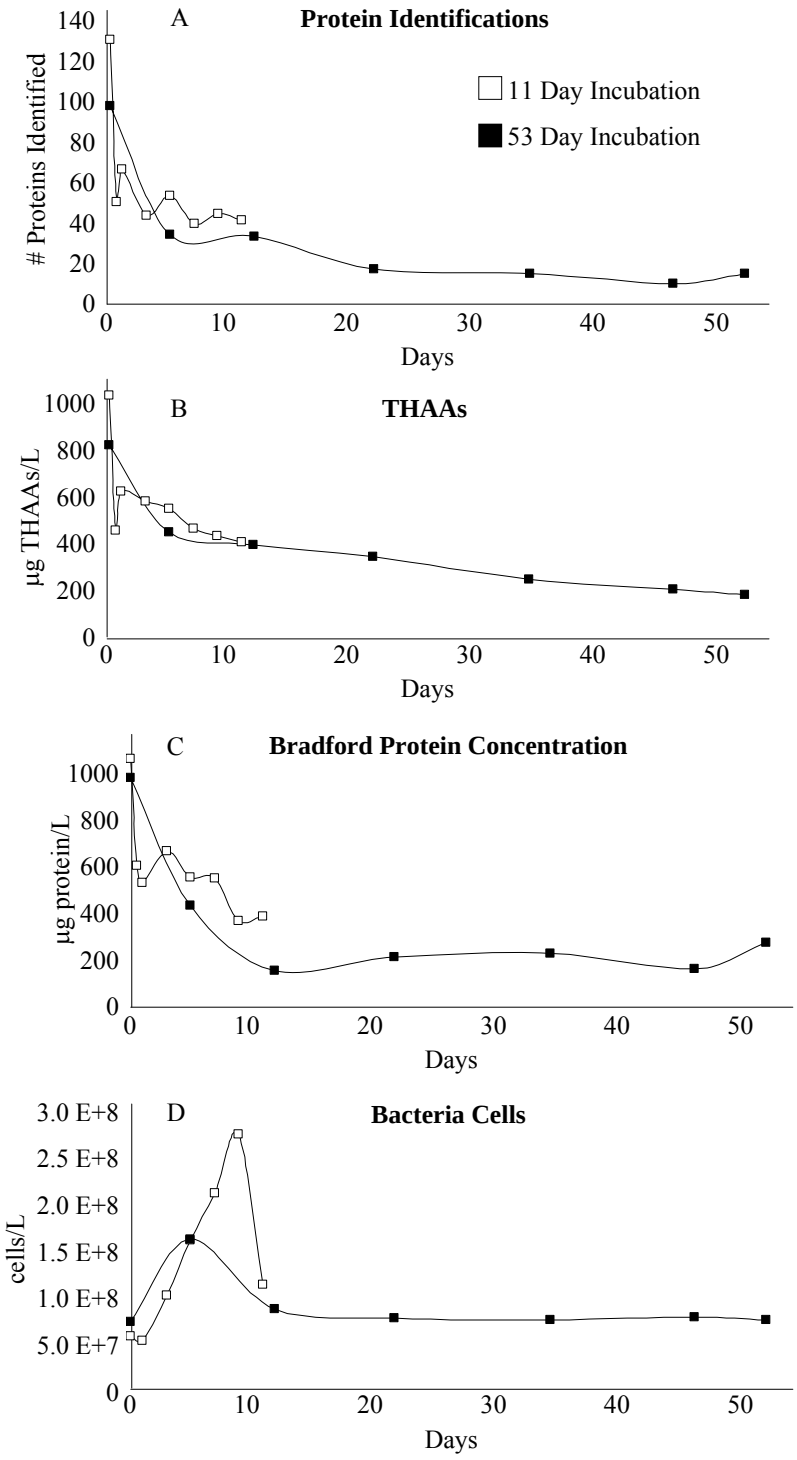

Figure 5-1. A) Total protein identifications; B) total hydrolysable amino acids (THAAs); C) total protein (Bradford assay); D) bacterial cell abundance of the course of the 11 day and 53 day incubations.

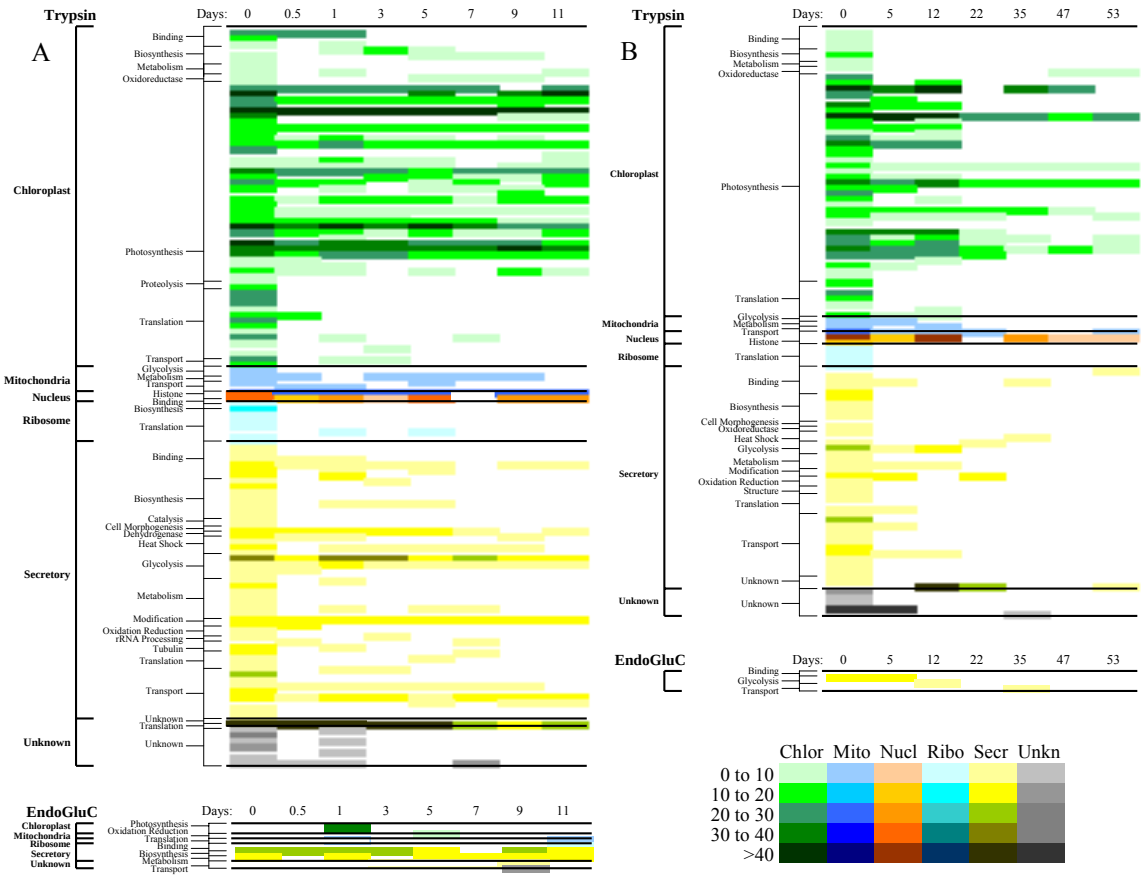

Figure 5-2. Identified proteins from each time point organized by compartment and function. Cellular compartments are represented by color and sequence coverage is represented by shade of each compartment color: A) 11 day trypsin and EndoGluC identifications; B) 53 day trypsin and EndoGluC identifications. Unknown represents proteins of uncharacterized function or cellular compartment.

the combined proportion of biosynthesis, glycolysis, metabolism, and translation proteins dropped from 38% and 27% at day 0 to 15% and 0% percent at the final time points of the 11 and 53 day incubations respectively.

The use of PNGase F resulted in the identification of 2 modified peptides in the 11 day incubation and 3 modified peptides in the 53 day incubation with the consensus sequence Asn-Xxx-Ser/Thr (Table 5-2). All modified peptides were identified in the first 5 days of both incubations. The unmodified tryptic version of the peptide from ATP synthase CF0 B chain subunit I was observed in the same PNGase F + trypsin digest as the modified form at day 5 of the 53 day incubation (Table 5-2). Inspection of the MS^2 fragmentation spectra of each peptide showed the mass change on b-ions that contained the altered asparagine for the modified peptide compared to the unmodified peptide (Figure 5-3 A, B). The unmodified versions of other apparent modified peptides were not identified in the same PNGase F + trypsin digests, however, the MS^2 fragmentation spectra still showed the same mass change on b-ions and y-ions containing the modified asparagine.

Few additional proteins were identified with digestion by EndoGluC with 8 additional identifications in the 11 day incubation, and 3 additional identifications in the 53 day incubation using EndoGluC (Figure 5-2 A, B). The largest increase in the number of identifications by using EndoGluC represented 4 additional identifications at days 1, 5, 9, and 11 in the 11 day incubation. Secretory proteins made up the majority of additional protein IDs using EndoGluC. There were no common protein IDs made using both PNGase F and EndoGluC that were not already identified using trypsin alone.

Table 5-2. Peptides identified with PNGase F + trypsin. These peptides contain the 1 Da mass increase on asparagine resulting in the 115 Da aspartic acid representing cleavage of a glycan chain. Also shown, the protein from which the peptides were identified and the day of the incubation during which the protein was identified.

11 Day Incubation

Protein	Peptide(s)	Day
RuBisCO large subunit	IHYLGDDVVLQFGGGTIGHPDGIQAGATAN[115]R	1
Predicted Protein	DLAEIWDN[115]SSPVIVQGGSLR	1

53 Day Incubation

Protein	Peptide(s)	Day
RuBisCO large subunit	TALDLWKDISFN[115]YTSTDTADFAETATANR	0
ATP synthase CF0 B chain subunit I	ALIN[115]ETIQKLEGDLL	5
Predicted Protein	QVVELYTEDGLDRPFFAIVETPGSGN[115]VVR	5

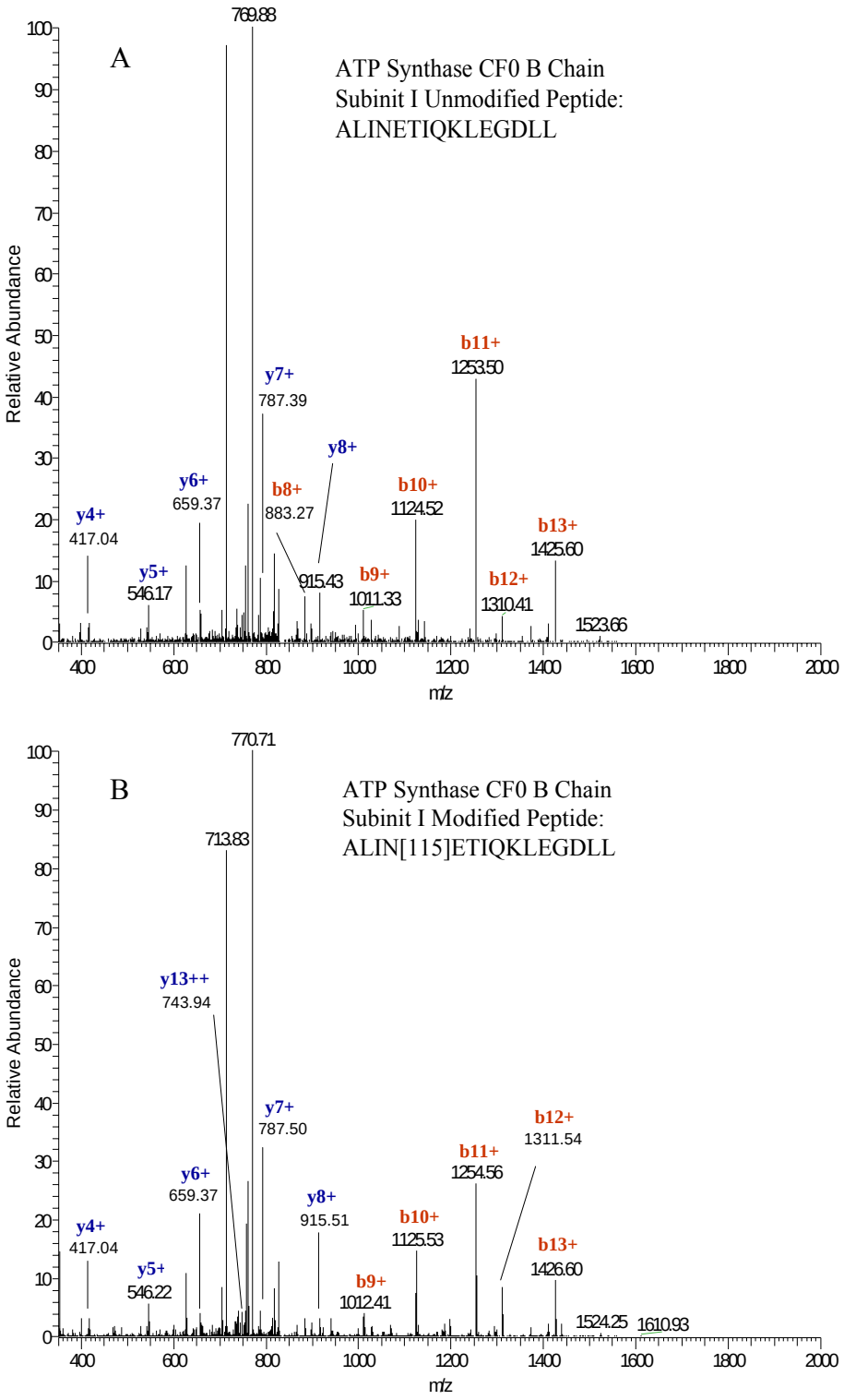

Figure 5-3. MS^2 spectra of ATP Synthase CF0 B chain subunit I peptide ALINETIQKLEGDLL showing b- and y-ions of the (A) unmodified and (B) modified peptide.

Discussion

The survival of algal proteins over 53 days of microbial recycling shows that a fraction of protein material remains identifiable long enough to potentially be exported from the marine water column to sediments. Changes to the suite of identifiable algal proteins appear to follow the path of proteome changes within the cells occurring first and followed by microbial recycling. After day 0 in the 11 day incubation the drop in the proportion of biosynthesis, glycolysis, metabolism, and translation proteins may represent the shutdown of cellular activity among algal cells. These changes take place in the first five days of both incubations, before microbial cell population peaks at day 9 of the 11 day incubation which is consistent with diatom cell activity timelines in darkness observed by Harvey et al., 1995. These trends in the changes to protein distribution appear to continue with microbial recycling as the proportion of biosynthesis, glycolysis, metabolism, and translation proteins continue to decline up to day 22 and are then no longer detectable in the 53 day incubation. Abundant photosynthetic proteins persisted and remained detectable after 53 days of recycling (Figure 5-2). At a sinking rate of up to or greater than 100 m day^{-1} (Smetacek, 1985), 53 days is more than enough time for some fraction of algal proteins from a diatom dominated system to transit down the water column as detritus to Bering Sea shelf and basin sediments as observed by Moore et al. (in review). This suggests that other diatom dominated ecosystems have the potential for export of identifiable protein from bloom to sediment.

At early time points in both incubations protein IDs, THAAs, and total protein decrease as bacterial cells increase, then this decrease is slower after bacterial cells return to original levels (Figure 5-1). Grazing by bacterivores, such as flagellates, protists, and

viral lysis (Steward et al., 1996; Vaque et al., 2008), may have limited bacterial recycling of protein material causing bacteria cell numbers to decrease and loss of identifiable protein, THAAs, and total protein to occur at a slower rate after 9 days. Despite the -1^{o}C ambient incubation temperature, bacterial cells were able to proliferate and appear to be primary recyclers of protein material. We can estimate the contribution of the bacterial fraction using cell numbers of observed bacteria and the average protein content per cell of 24 fg cell^{-1} (Zubkov et al., 1999). The bacterial fraction of total protein and THAAs was highest at or near the peak bacteria cell abundance in both the 11 and 53 day incubations (Table 5-1), but was never greater than two percent. The simple database used primarily to observe algal proteins, and the potentially high diversity of the microbial community in the incubation may explain why identifiable proteins were dominated by diatom sources with few prokaryotic proteins identified.

The number of peptides identified for each protein generally decreased through the course of the 53 day incubation (Figure 5-4). Of the 11 proteins identified at day 53 of the long incubation, 7 were located in the chloroplast. Five and three peptides were identified from the RuBisCO large and small subunits respectively at 53 days, contrary to previous findings that RuBisCO was degraded more rapidly and not identifiable as long as other chloroplast proteins in a lab based degradation of *T. pseudonana* (Nunn et al., 2010). Two fucoxanthin chlorophyll binding proteins (FCPs) were identified throughout the 53 day incubation in agreement that these proteins remain identifiable for extended periods in lab degradation (Nunn et al., 2010) and Bering Sea shelf (139 m) and basin (3490 m) sediments (Moore et al., in review).

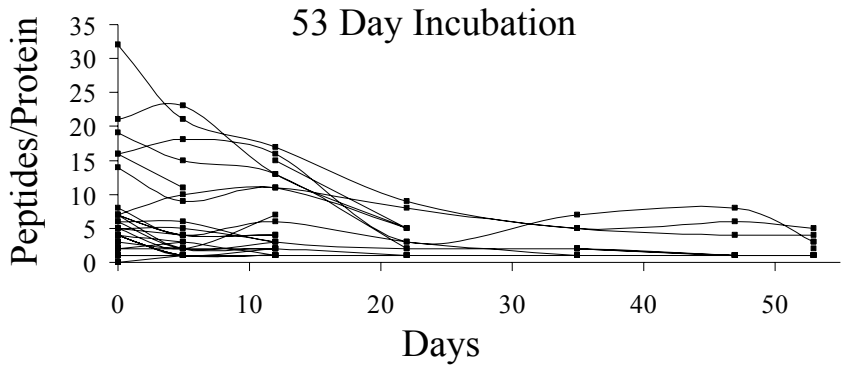

Figure 5-4. Plot of the number of peptides identified from each protein during the 53 day incubation.

The ratios of THAA/OC and THAA-N/PN have been observed as traditional proxies for degradation status of organic matter (Cowie and Hedges, 1994 and references therein). Plotting the number of IDs against the ratio of THAA/OC and THAA-N/PN (Figure 5-5), one can observe that all of the 11 day incubation time points and the three earliest time points of the 53 day incubation (days 0, 5, 12) have greater numbers of protein IDs and higher ratios of THAA/OC and THAA-N/PN than the four latest time points of the 53 day incubation (days 22, 35, 47, 53), which cluster together more closely. The change suggests that after initial rapid degradation, identifiable protein material degrades very slowly for over a month. This longevity of identifiable protein may allow proteomic characterization of material measured as THAAs throughout sediments of the global ocean.

Non-metric multidimensional scaling (theory and applications described in Borg and Groenen, 2005), performed using R Statistical Software, was used in order to group incubation time points with water column particles and surface sediments from a previous study (Moore et al., in review) based on the distribution of proteins in each sample. Potential similarities between incubation time points and water particles or sediments could allow approximate time-frames to be assigned to samples in the field (Figure 5-6). Early incubation time points (days 0 to 11) cluster closely together as do later incubation time points (days 22 to 53). Chl max particles and sinking sediment trap particles are positioned more closely to initial incubation time points than later time points for both the 11 and 53 day incubations. Post bloom shelf surface sediment is positioned closer to early incubation time points while post bloom basin sediment and over wintered shelf sediment are closer to later incubation time points. The observation that 50 m and 100 m

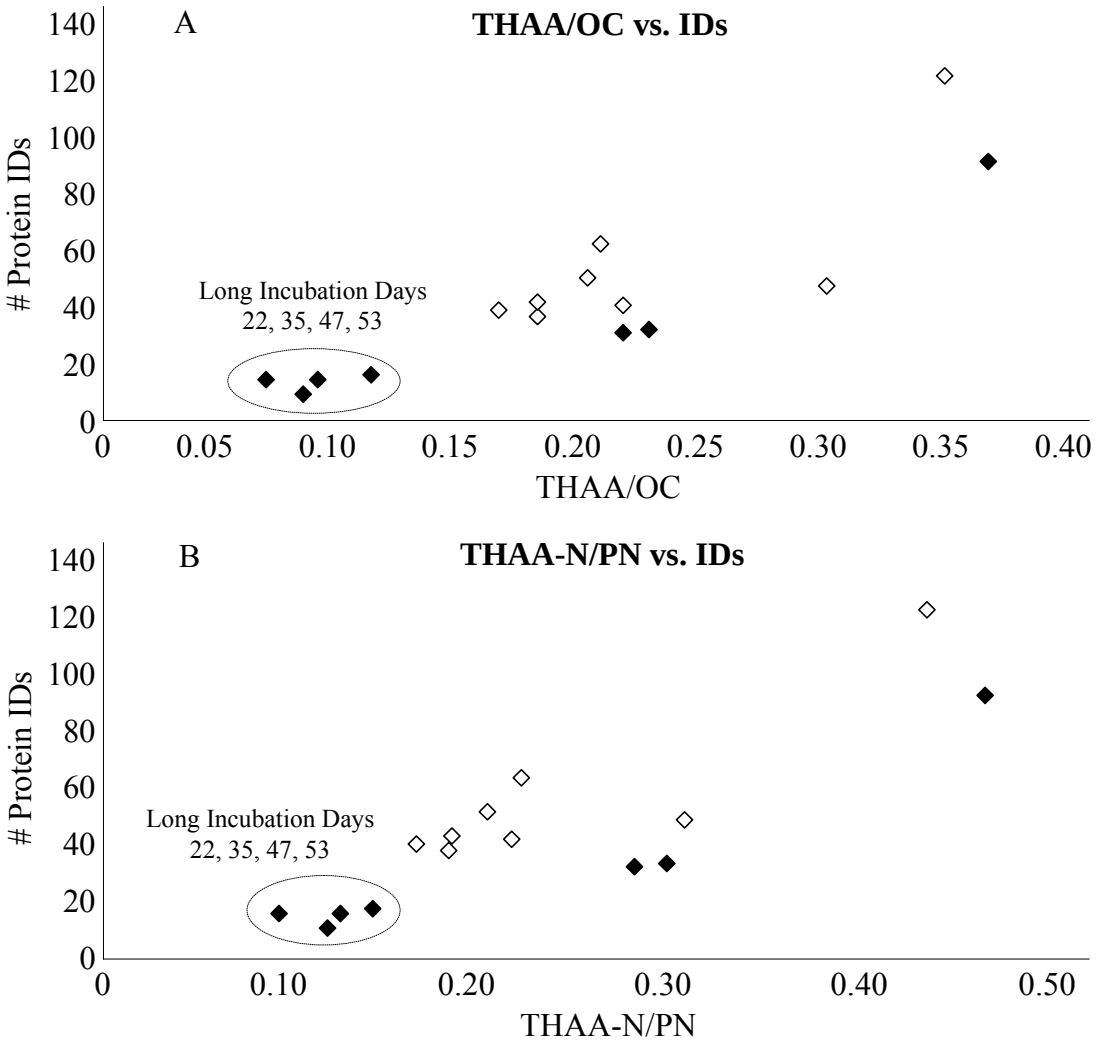

Figure 5-5. Plots of total protein identifications vs. A) THAA/OC and; B) THAA-N/PN.

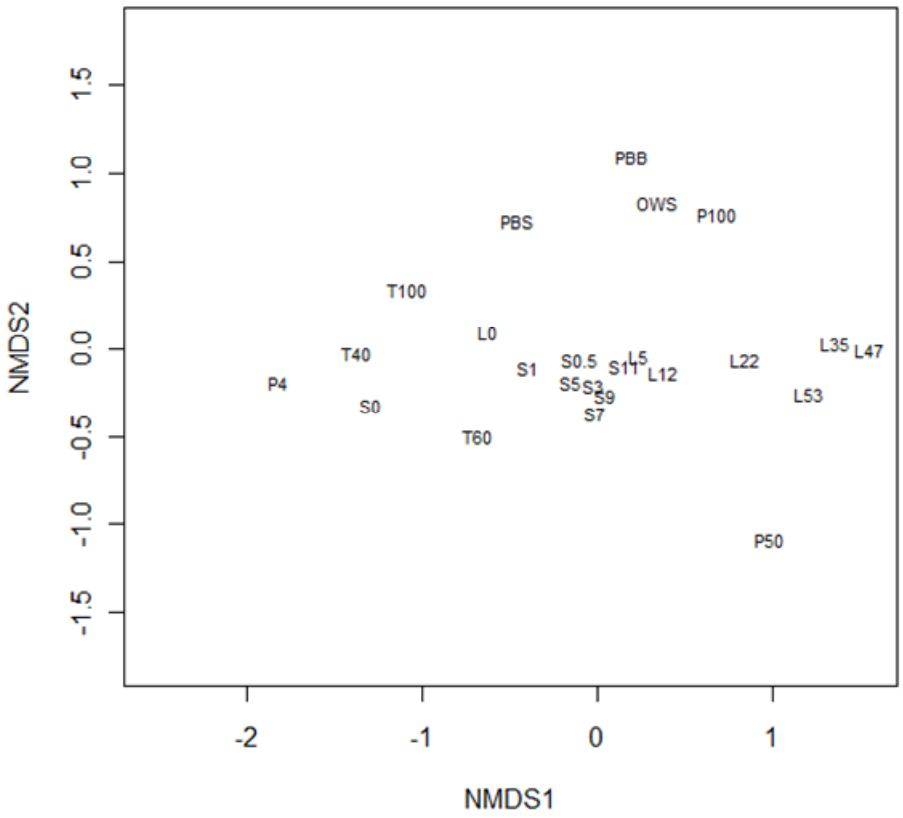

Figure 5-6. Non-metric multidimensional scaling ordination based on the protein distribution of Bering Sea water column and sediment samples (Moore et al., in review) and the time points of the 11 and 53 day incubations: P4 = Chl Max suspended particles; P50 = 50 m suspended particles; P100 = 100 m suspended particles; T40 = 40 m sediment trap; T60 = 60 m sediment trap; T100 = 100 m sediment trap; PBS = Post Bloom Shelf sediment; PBB = Post Bloom Basin sediment; OWS = Over Wintered Shelf sediment; 11 Day incubation days 0, 0.5 1, 3, 5, 7, 9, 11 = S0, S0.5, S1, S3, S5, S7, S9, S11; 53 Day incubation days 0, 5, 12, 22, 35, 47, 53 = L0, L5, L12, L22, L35, L47, L53.

suspended particles are positioned slightly closer to later time points than sediments illustrates the importance of export to sediments towards protein preservation

Non-enzymatic glycation is a major cause of spontaneous damage to cellular and extracellular proteins in physiological systems (Ahmed and Thornalley, 2007). Various mechanisms have been observed in higher plants (Kim and Kim, 2003; Bechtold et al., 2009; Sultana et al., 2009;) and the microalga *Chlorella zofingiensis* (Sun et al., 2011) to prevent and correct protein glycation. Glycation of RuBisCO has been shown to decrease the enzyme's activity, and increase its susceptibility to proteases (Yamauchi et al., 2002). Thus, the potential glycation of RuBisCO and other proteins (Table 5-2) in the early stages of recycling may represent protein turnover within the cell. Glycation has also been hypothesized as a mechanism for protein preservation and the formation of geopolymers (Collins et al., 1992; Fogel and Tuross, 1999; Burdige, 2007). Mass spectral analysis with InsPecT (Tanner et al., 2005) was used to identify recurring modification masses on precursor ions revealing potential sugar modifications to proteins from degraded phytoplankton (Nunn et al., 2010). While difficulties arise as observed with distinguishing modifications with isotopic mass shifts, identifying protein modifications by proteomic methods to identify mass changes will help enhance the ability to identify degraded proteins in environmental samples.

Conclusions

The persistence of algal proteins during in situ degradation experiments indicates that some fraction of this material can survive water column recycling to be transported to marine sediments. Initial changes to algal cell proteomes and bacterial recycling appear to result in rapid loss of identifiable proteins followed by slow degradation of material after 22 days. The slowing of degradation after 22 days was observed in identifiable proteins, total protein, THAA/OC, and THAA-N/PN. As fewer proteins were identified, fewer peptides per protein were identified as well. The observation of modified peptides with the addition of PNGase F suggests that additional proteins may be identified by accounting for protein modification in degraded samples.

Chapter 6: Conclusions and Implications

Marine proteomics: Expanded characterization of organic matter

The work presented represents a major step forward in characterizing marine organic nitrogen from the water column to sediment. Historically only bulk elemental and amino acid analyses were available for examining the protein component of marine particles and sediment (Nunn and Timperman, 2007). The perceived lability of protein suggested that observed marine hydrolysable amino acids did not represent identifiable proteins or peptides. However, this work shows that some fraction of environmental protein can retain identifiable amino acid sequences. Particles and sediments where hydrolysable amino acids (Lee and Cronin, 1982; Hedges, 1991; Benner et al., 1992; Keil et al., 1994; McCarthy et al., 1998; Horiuchi et al., 2004; Wakeham et al., 2010) and amide linked nitrogen (McCarthy et al., 1997; Knicker, 2000) have been measured, throughout the ocean, can now be considered for proteomic analysis to identify amino acid source. This will expand the knowledge of carbon and nitrogen export in the ocean as amino acids are important contributors to both pools (Wakeham et al., 1997; Keil, 1999; Nunn, 2004).

The finding that algal proteins from a diatom dominated system can be tracked down the water column to surface sediments of the Bering Sea continental shelf and deeper basin shows that diatom bloom material can undergo transport to sediments. This is in agreement with previous findings that water column microbial recycling is reduced during periods of high carbon flux to sediments (Abramson et al., 2010; Wakeham et al., 2010), or in cold waters (Pomeroy and Deibel, 1986), and that diatom blooms can lead to efficient transfer of particles through the mesopelagic (Martin et al., 2011). As diatoms

are responsible for up to 40% of marine primary production (Granum et al., 2005), they are likely one of the most important conduits for carbon and nitrogen export in the form of identifiable protein in many systems around the globe.

Diatoms are not the only group of phytoplankton whose cells aggregate and sink, leading to particle export. It has been estimated that *Phaeocystis antarctica*, a colonial haptophyte, contributes 30% of carbon export in the Southern Ocean (>60 degrees S) (Wang and Moore, 2011). Recent observations have also indicated that nano- and pico-plankton may represent a significant contribution to the total particulate organic carbon (POC) export via formation of aggregates in an oligotrophic gyre of the Sargasso Sea (Lomas and Moran, 2011). Annual global estimates from SeaWiFS satellite observations from 1998 to 2007 by Uitz et al (2010) amount to 15 Gt C yr^{-1} (32% of total), 20 Gt C yr^{-1} (44%), and 11 Gt C yr^{-1} (24%) for micro- (mostly diatoms), nano- (e.g., prymnesiophytes), and picophytoplankton (e.g., prokaryotes) respectively. The range of ecosystems from which phytoplankton export could be characterized and the increasing number of phytoplankton genomes that are being sequenced shows that proteomic analysis of primary production down the water column to surface sediments is not just limited to diatom dominated regions.

The information embedded in each protein's amino acid sequence holds the functional role which that protein carries out in the cell. Thus, identifying the suite of proteins in a microbial community can reveal the biogeochemical activities of an entire population (ex: carbon or nitrogen fixation; sulfur reduction; ion transport; pH homeostasis). Conserved peptide sequences observed by Powell et al. (2005) identified membrane/envelope proteins and enzymes providing insight into the sources and

production mechanisms for dissolved organic matter. Shifts in nutrient utilization and energy transduction have also been observed from a low-nutrient gyre to highly productive coastal upwelling region (Morris et al., 2010). We observed trends in this study showing that the proportion of chloroplast and transmembrane proteins increased down the water column to sediments revealing potential mechanisms that influence protein preservation. While further work is needed to develop microbial databases in marine sediments, this work has shown the importance of algal inputs to polar continental shelf sediments, and will help pave the way for further studies on nitrogen storage and remineralization.

New methods to observe protein function and fate in sediments and soils

The sediment gel electrophoretic extraction methods developed in this study were successful purifying both protein standard (BSA) and environmental proteins from marine sediment. This method can be applied to samples from a range of environments, as the combination of protein solubilizing agents and electrophoresis current appear to liberate proteins from mineral sorption and organic matter interaction/interference, obstacles which are universal among sediment and soil samples. This study also demonstrates that complex protein databases, such as the NCBI or GOS which provide more potential protein sources, do not necessarily translate into a greater number of confident protein identifications than the Thaps database (includes proteomes of *T. pseudonana, P. marinus, P. ubique*) for Bering Sea sediment, due to inherent issues with searching larger databases. As database size increases the potential for false positives also increases, which reduces the number of statistically significant identifications.

Functional-level information is retained, despite the organism with which the protein is associated using different databases, which demonstrates that identifying proteins from mixed (often unknown) communities can be accomplished at the protein function-level, although determining and/or targeting the specific species the protein originated from remains difficult. These issues will diminish as database quality improves and more relevant genomes are sequenced.

Protein identification in sediments and soil is of great interest to the scientific community as proteins contain key genetic information to microbial community phylogeny, function and activity (Schulze et al., 2005, Bastida et al., 2009). Furthermore, preserved proteins could potentially be used towards phylogenetic and biogeochemical reconstruction (Ostrom et al., 2000 and references therein), or forensic and archaeological identification (Loy and Hardy, 1992; Tuross et al., 1996). These applications of protein identification can be carried out with greater phylogenetic resolution than phosopholipid fatty acids (Zelles, 1997) and more information regarding organism activity than PCR amplification and sequencing of conserved rRNA sequences. Lipid and DNA techniques have largely been restricted to determine taxonomic relationships of individual species (Harris et al., 2002) and description of seasonal variations in soil microbial communities (Lipson et al., 2002). The methods from this study may help provide new avenues to scientists long seeking to identify proteins from complex matrices.

Protein database searching limitations and tailoring shotgun proteomic studies to a specific study system

Protein sequence database searching is a crucial step in using proteomic mass spectrometry (MS) to assign source and function to proteins in environmental samples. The wide availability of completely sequenced genomes from a number of species has greatly expanded the applicability of MS based protein identifications, as the requirement for *de novo* sequencing has been usurped by simple correlation of measured data versus theoretical data from sequence databases (Aebersold and Mann, 2003). However, mass spectrometers are limited in their dynamic range, which restricts the capability to detect very low abundance analytes in biological samples with large dynamic range (Prakash et al., 2007). Furthermore, in a protein sequence database search, the peptide and protein sequences that can be identified in a sample are limited to the sequences that are in the database being searched. The limitations of mass spectrometers and database searches should be taken into account when analyzing environmental samples to optimize results.

Protein sequence databases are only useful for known proteins. If a protein sequence is not documented it will not be found during a database search. Furthermore, only a small percentage of known proteins have been studied experimentally (Lubec and Afjehi-Sadat, 2007). However, many protein sequences are conserved among different organisms allowing undocumented proteins to be identified as a protein from another organism. In this case, since protein function is a result of amino acid sequence, the protein's function is likely assigned correctly despite potentially incorrect source organism identification. This was observed in our database comparison, where 95% functional agreement was observed between peptides that were identified from searches

of multiple databases. Sequence modifications compound the issue of unknown protein sequences resulting in non-detections as most databases rely on mass spectral data matching only the known unmodified protein amino sequences. New methods such as spectral networking (Bandeira et al., 2007) have been developed to address the issues of posttranslational modifications, but they are better suited for single cell cultures than diverse environmental samples.

Because of the great diversity of marine microbial communities it is tempting to include as many protein sequences as possible in a protein database to maximize the potential number of proteins that could be identified in a given sample. Large protein databases like the translated DNA from Global Ocean Sampling effort led by Craig Venter (Yooseph et al., 2007), and the National Center for Biotechnology Information Reference Sequence non redundant database (Pruitt et al., 2007) attractively allow a researcher to potentially identify a very large number of proteins from a complex sample. However, as database size increases, the potential for false positive protein identification also increases. This indicates that there is a balancing act between constructing a database that includes relevant protein sequences to address one's research goals while limiting database size to avoid false positives. For example, a database used for characterizing a productive coastal upwelling region is likely to be much different than for an oligotrophic gyre.

Database searching software such as SEQUEST, MASCOT, or X!Tandem, also have limitations that hinder protein identification. Spectral quality, based on charge state differentiation, total signal intensity, and signal to noise estimates, is of the utmost importance in protein identification (Lubec and Afjehi-Sadat, 2007). This is especially

true when working with complex samples like marine particles or sediments which invariably include interference materials. Various efforts have been made to remove poor quality spectra using pre-filtering rules (Tabb et al., 2000) or algorithms (Bern et al., 2004), however such actions have the potential to discard useful good quality spectra. Database quality is vital to protein identification as well since searching software relies completely the database for protein identification. Database errors have been observed as a consistent problem in both academia and the biotechnology industry (Hadley, 2003). Researchers should strive to use high quality mass spectra and curated databases in order minimize protein identification error.

Confounding the issues of spectra and database quality, searching software may not be well equipped to handle searching mass spectra against large databases. As seen from this study in the discussion of xcorr and ΔCorr scoring, in some cases PeptideProphet (Keller et al., 2002: used in conjunction with SEQUEST) can reject spectra that match well with multiple peptide sequences. This was evident in the fact that more proteins were identified using a smaller protein sequence database containing the proteomes of three organisms vs. large databases containing millions of protein sequences. As a result this shows that SEQUEST and PeptideProphet, or other database searching software which use similar statistical scoring to remove perceived low quality spectra, are biased towards smaller databases. This may require the researcher to refine larger databases to prevent searching software from rejecting potentially useful spectra.

In order to successfully apply shotgun proteomics to marine particles or sediments it is essential to know the ecosystem being studied: who are the big players in the microbial community and which ones are being studied? This information can then be

used properly prepare samples and use the appropriate protein sequence database to suit the goals of the research. The dynamic range limitation of mass spectrometers biases the results towards high abundance proteins in a given sample. That means that the most abundant proteins from the most abundant organisms have a higher probability of being detected than proteins from low abundance organisms. Water column particles in the Bering Sea, a highly productive diatom dominated system, are expected to contain high concentrations of diatom proteins. Thus large volumes of water were not needed to adequately sample diatom material and address our goals of tracking primary production down the water column to sediments.

To identify proteins in samples expected to contain diatoms, the proteome of *T. pseudonana*, a well characterized marine diatom, was included in a small protein database containing the proteomes of three organisms. If the research goals were to characterize the protein distribution of the bacterial community during a Bering Sea diatom bloom, it would be necessary to filter out the diatom cells using an appropriate filter pore size to remove material that may obscure detection of the desired bacterial proteins, and then filter very large volumes of the pre-filtered water on a much smaller filter pore size to boost the bacterial biomass and increase the likelihood of detection. The resulting mass spectra would then be searched against a database containing the proteomes of microbes expected to be associated with a diatom bloom in the Bering Sea. In the case of the Bering Sea, the biomass of diatoms during the spring phytoplankton may be so much larger than that of bacteria, that the two organism size fractions must be analyzed exclusively to overcome the challenge of dynamic range in mass spectrometry.

The issues of organism diversity and low biomass will be common obstacles in using proteomic mass spectrometry to characterize marine microbial communities. Large scale DNA sequencing efforts like the Global Ocean Sequencing project (Yooseph et al., 2007) have greatly increased the number of known marine protein sequences, but many unknown protein sequences still remain. Because microbes are so diverse, individual bacteria or archaea species and their proteins likely make up small portions of the total microbial biomass in a given sample. Thus, individual proteins are below detection limits (Figure 6-1). In order identify such low abundance proteins, efforts must be made to isolate and amplify targeted organism proteins. Morris et al. (2010) successfully observed thousands of microbial cell membrane peptides from the South Atlantic by filtering large volumes of seawater (100-200 L), followed by size fractionation and membrane enrichment. Specialized sample manipulation will continue to be important for investigating microbial protein distribution and function using proteomic mass spectrometry and database searching.

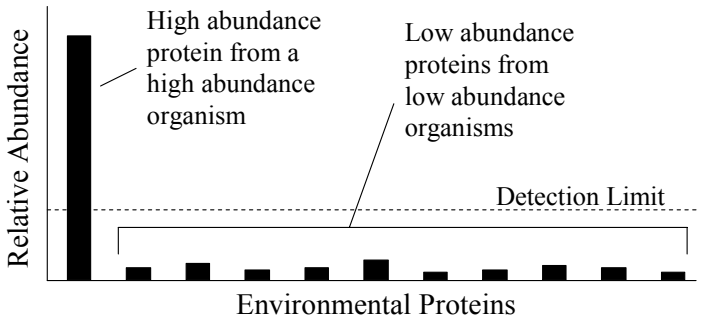

Figure 6-1. Depiction of environmental protein abundances affecting their identification based on instrument detection limits. Each bar represents the relative abundance of an individual hypothetical protein.

Shipboard incubations and protein longevity

During the course of the 11 and 53 day shipboard incubations initial rapid loss of identifiable proteins, THAAs, and total protein coincided with the increase of bacterial cell concentrations. When bacterial cell numbers returned to their initial levels, loss of identifiable proteins, THAAs, and total protein proceeded much more slowly. This was particularly evident after 22 days in the 53 day incubation. Protein material that survives the potential cellular restructuring during bloom termination followed by microbial recycling may be an important component of marine refractory organic nitrogen. The time scales of the observable suite of proteins from primary production will be important for estimating the lifetime of exportable carbon and nitrogen from the photic zone.

Marine proteomics as a an approach to describe carbon and nitrogen cycling under climate change conditions

The observed transport of diatom protein from bloom to sediment in these studies raise issues for the future organic matter export in diatom dominated regions. In the Bering Sea, in situ incubation studies were conducted to simulate global warming conditions on algal and microbial communities and observe how they respond compared to current conditions (Hare et al., 2007). Global warming conditions were simulated by increasing the incubation temperature and CO_2 partial pressure. It was observed that while primary production increased, the phytoplankton community changed from diatom to nano-plankton dominated. The effects of increased pCO2 and temperature on phytoplankton were also carried out in the North Atlantic during the spring bloom (Feng

et al., 2009). Again, a phytoplankton community shift from diatoms to nanoplankton was observed.

A shift in phytoplankton community could have major implications on the carbon cycle as the dense silica cell walls of diatoms make them much more efficient carbon export vectors then nano-plankton species (Dunne et al., 2005; Ragueneau, et al., 2006; Miki et al., 2009). If this observed trend in the Bering Sea and North Atlantic took place in other diatom dominated systems, it in turn would result in less export of organic carbon and nitrogen, a weakening of the global biological pump, and a positive feedback in global warming atmospheric carbon. There are different findings with regard to how phytoplankton will respond increased pCO2 and ocean acidification (Iglesias-Rodriguez et al., 2008; Guinotte and Fabry, 2008; Shi et al., 2010; Beaufort et al., 2011), including changes to the transcription of calcium and bicarbonate transporters (Richier et al., 2011). The work from this thesis shows that proteomics can be a direct method of observing phytoplankton physiological responses and reconstructing water column inputs of organic carbon and nitrogen to sediments in response to environmental change.

Appendices

APPENDIX 3-1

Summary list of all proteins identified in each sample database search. Includes identified species, biological function, cellular compartment (Comp): C = Chloroplast; S = Secretory; M = Mitochondria; N = Nucleus; U = Uncharacterized Compartment, molecular weight (MW in Da), isoelectric point (pI), percent sequence coverage (SC).

Slurry Tube Gel – Thaps Database

Protein	Annotation		Species	Function	Comp	MW	pI	SC
30S ribosomal protein S1	jgi	15259	*T. pseudonana*	Binding, RNA	R	31770	4.6	3.8
30S ribosomal protein S2	gi	1184 11126	*T. pseudonana*	Translation	C	25635	9.3	11.9
30S ribosomal protein S5	gi	1184 11207	*T. pseudonana*	Translation	C	19297	10.3	11.7
30S ribosomal protein S6	gi	1184 11222	*T. pseudonana*	Translation	C	12043	9.8	10.7
30S ribosomal protein S7	gi	1184 11217	*T. pseudonana*	Translation	C	17730	10.5	21.2
3-deoxy-7- phospho-heptulonate synthase	jgi	2790	*T. pseudonana*	Biosynthesis, amino acid	C	53939	6.0	2.5
40S ribosomal protein S11	jgi	22535	*T. pseudonana*	Translation	R	19118	10.3	8.5
40S ribosomal protein S23	jgi	28209	*T. pseudonana*	Translation	R	15734	10.5	12.6
40S ribosomal protein SA	jgi	21871	*T. pseudonana*	Translation	R	27261	5.9	10.3
50S ribosomal protein L11	gi	1184 11123	*T. pseudonana*	Translation	C	14880	9.7	9.2
50S ribosomal protein L14	gi	1184 11201	*T. pseudonana*	Translation	C	13433	10.3	13.2
50S ribosomal protein L19	gi	1184 11099	*T. pseudonana*	Translation	C	13812	10.6	25.0
50S ribosomal protein L2	gi	1184 11193	*T. pseudonana*	Translation	C	30675	10.9	5.5
50S ribosomal protein L22	gi	1184 11196	*T. pseudonana*	Translation	C	12986	10.3	11.3
50S ribosomal protein L23	gi	1184 11192	*T. pseudonana*	Transferase	C	11152	9.9	10.2
50S ribosomal protein L24	gi	1184 11202	*T. pseudonana*	Biosynthesis, amino acid	C	8682	10.1	20.8
50S ribosomal protein L3	gi	1184 11190	*T. pseudonana*	Translation	C	22012	10.1	26.1
50S ribosomal protein L4	gi	1184 11191	*T. pseudonana*	Translation	C	24203	10.2	7.9
50S ribosomal protein L5	gi	1184 11203	*T. pseudonana*	Translation	C	27571	9.7	3.8
60 kDa chaperonin	gi	1184 11188	*T. pseudonana*	Metabolic	C	57361	5.2	3.2
6-phospho-gluconate dehydrogenase	jgi	33343	*T. pseudonana*	Pentose Phosphate Shunt	S	53348	5.6	9.8
Abnormal wing discs	jgi	6290	*T. pseudonana*	Diphosphate kinase	U	17236	5.5	13.8
Acetyl-CoA carboxylase	jgi	6770	*T. pseudonana*	Ligase	C	228295	5.0	3.2

Acidic ribosomal phosphoprotein P0	jgi	25812	*T. pseudonana*	Biogenesis	R	34116	4.8	3.7
Actin A	jgi	25772	*T. pseudonana*	Cytoskeleton	S	41791	5.0	4.8
Adenosine-triphosphatase	jgi	40156	*T. pseudonana*	Transport, proton	S	39935	7.6	10.6
Adenylate kinase	jgi	31809	*T. pseudonana*	Metabolic, nucleic acid	S	24598	6.3	4.4
ATP synthase CF0 B chain subunit I	gi	1184 11110	*T. pseudonana*	Photosynthesis	C	20029	9.8	8.4
ATP synthase CF0 B' chain subunit II	gi	1184 11109	*T. pseudonana*	Photosynthesis	C	17373	4.6	9.6
ATP synthase CF0 C chain subunit III	gi	1184 11108	*T. pseudonana*	Photosynthesis	C	8166	5.0	50.0
ATP synthase CF1 alpha chain	gi	1184 11112	*T. pseudonana*	Photosynthesis	C	53989	5.0	8.5
ATP synthase CF1 beta chain	gi	1184 11134	*T. pseudonana*	Photosynthesis	C	51143	4.7	30.2
ATP synthase CF1 delta chain	gi	1184 11111	*T. pseudonana*	Photosynthesis	C	21077	9.2	16.0
ATP/ADP translocator	jgi	39143	*T. pseudonana*	Transport	S	32254	9.4	9.9
ATP-dependent clp protease ATP-binding subunit	gi	1184 11220	*T. pseudonana*	Catalytic activity	C	102150	6.5	3.0
BiP	jgi	27656	*T. pseudonana*	Morphogenesis, Cell	C	70451	4.7	3.6
CbbX protein homolog	jgi	40193	*T. pseudonana*	Binding, ATP	C	35036	5.3	18.3
CDC48/ATPase	jgi	267952	*T. pseudonana*	Binding, ATP	S	89464	4.8	3.1
Cell division protein FtsH2	jgi	31930	*T. pseudonana*	Binding, Zn	S	61956	5.1	1.9
Cell division protein FtsH-like protein	gi	1184 11141	*T. pseudonana*	Photosynthesis	C	70206	5.1	6.9
CG11154-PA, isoform A	jgi	41256	*T. pseudonana*	Transport, proton	S	53388	5.1	9.4
Chloroplast light harvesting protein isoform 12, 18 kDa	jgi	33606	*T. pseudonana*	Photosynthesis	C	18463	4.6	7.6
Chloroplast light harvesting protein isoform 12, 26 kDa	jgi	270092	*T. pseudonana*	Photosynthesis	C	26078	5.5	6.2
Chloroplast light harvesting protein isoform 15	jgi	2845	*T. pseudonana*	Photosynthesis	C	21873	5.1	5.4
Chloroplast light harvesting protein isoform 5	jgi	32723	*T. pseudonana*	Photosynthesis	C	19175	5.2	10.6
Citrate synthase, mitochondrial precursor	jgi	11411	*T. pseudonana*	Transferase	M	52269	6.2	2.3
Coatomer protein complex, subunit gamma 2	jgi	269663	*T. pseudonana*	Transport, protein	S	100270	5.1	1.1
Cyclophilin	jgi	29244	*T. pseudonana*	Folding, Protein	S	20899	6.9	9.2
Cytochrome b559 alpha chain	gi	1184 11160	*T. pseudonana*	Photosynthesis	C	9514	5.6	21.4
Cytochrome b6	gi	1184 11154	*T. pseudonana*	Photosynthesis	C	23906	9.2	6.0
Cytochrome f	gi	1184 11137	*T. pseudonana*	Photosynthesis	C	33988	8.2	4.1
Elongation factor 2	jgi	269148	*T. pseudonana*	GTPase	S	91887	6.0	8.7
Elongation factor alpha-like protein	jgi	41829	*T. pseudonana*	GTPase	S	49969	8.7	16.8
Enoyl-acyl carrier reductase	jgi	32860	*T. pseudonana*	Oxidation reduction	S	32813	5.1	3.5
Eukaryotic translation initiation	jgi	9716	*T. pseudonana*	Binding, DNA	S	42405	5.6	11.9

factor 4A2 isoform 2								
Ferredoxin component	jgi	29842	*T. pseudonana*	Oxidoreductase	C	18511	8.9	11.7
Fructose-1,6-bisphosphate aldolase precursor	jgi	428	*T. pseudonana*	Glycolysis	S	39810	4.8	8.1
Fucoxanthin chlorophyl a/c protein, 21 kDa	jgi	38667	*T. pseudonana*	Photosynthesis	C	21807	4.8	17.6
Fucoxanthin chlorophyll a /c protein, 20 kDa	jgi	38494	*T. pseudonana*	Photosynthesis	C	20354	4.5	11.6
Fucoxanthin chlorophyll a /c protein, 21 kDa	jgi	42962	*T. pseudonana*	Photosynthesis	C	21515	5.1	11.0
Fucoxanthin chlorophyll a/c binding protein, 22 kDa	jgi	264921	*T. pseudonana*	Photosynthesis	C	22205	4.6	8.6
Fucoxanthin chlorophyll a/c binding protein, 22.6 kDa	jgi	268127	*T. pseudonana*	Photosynthesis	C	22628	4.8	17.1
GDP-mannose 3,5-epimerase	jgi	41548	*T. pseudonana*	Coenzyme	S	40707	5.1	3.3
GDP-mannose dehydratase	jgi	40586	*T. pseudonana*	Coenzyme	S	40412	5.9	4.2
Geranyl-geranyl reductase	jgi	10234	*T. pseudonana*	Metabolic, Aromatic Compound	C	47230	5.9	10.3
Glucose-6-phosphate dehydrogenase	jgi	34514	*T. pseudonana*	Metabolic, glucose	S	57149	7.6	2.2
Glyceraldehyde-3-phosphate dehydrogenase	jgi	28334	*T. pseudonana*	Glycolysis	M	36574	6.1	10.8
Glyceraldehyde-3-phosphate dehydrogenase precursor	jgi	31383	*T. pseudonana*	Glycolysis	S	39587	5.3	14.1
Glycolaldehy-detransferase	jgi	21175	*T. pseudonana*	Transport	M	71708	5.0	2.3
Heat shock protein 70	jgi	269120	*T. pseudonana*	Heat shock	S	71187	4.8	8.1
Heat shock protein Hsp90	jgi	6285	*T. pseudonana*	Heat shock	S	80242	4.7	2.0
Histone H2A.1	jgi	19793	*T. pseudonana*	Binding, DNA	N	13053	10.4	7.3
Histone H4	jgi	3184	*T. pseudonana*	Binding, DNA	N	11384	11.5	35.9
HLA-B associated transcript 1	jgi	269556	*T. pseudonana*	Hydrolase	S	49087	5.2	2.8
Hsp70-type chaperone	gi	1184 11189	*T. pseudonana*	Transcription	C	65339	4.8	2.3
HSP90-like protein	jgi	22766	*T. pseudonana*	Heat shock	S	80966	4.7	2.0
Hypothetical protein CBG08717	jgi	269322	*T. pseudonana*	Transport, Proton	S	58009	5.8	13.8
Hypothetical protein CdifQ_02003487	jgi	36462	*T. pseudonana*	Metabollic	U	21340	4.8	6.7
Hypothetical protein FG01081.1	jgi	25949	*T. pseudonana*	Translation	R	20124	9.8	15.3
Hypothetical Protein No BLAST result	jgi	23918	*T. pseudonana*	n.a.	U	15261	10.0	11.6
Hypothetical Protein No BLAST result	jgi	11169	*T. pseudonana*	n.a.	S	31192	4.8	14.0
Hypothetical protein SPBC29A3.04	jgi	269961	*T. pseudonana*	n.a.	C	28946	10.5	3.8
Isocitrate dehydrogenase NADP-dependent, monomeric type	jgi	1456	*T. pseudonana*	TCA Cycle	C	73809	5.7	2.6

Ketol-acid reductoisomerase	jgi	23228	*T. pseudonana*	Biosynthesis, Amino Acid	S	58240	5.1	3.2
L4/L1	jgi	22610	*T. pseudonana*	Translation	R	40991	10.3	7.7
Nucleoside diphosphate kinase	jgi	31091	*T. pseudonana*	Diphosphate kinase	S	16917	8.3	11.3
Oxygen-evolving enhancer protein 1 precursor	jgi	34830	*T. pseudonana*	Photosynthesis	S	29136	5.2	20.4
Phosphoglycerate kinase precursor	jgi	35712	*T. pseudonana*	Glycolysis	S	42256	5.0	10.5
Photosystem I ferredoxin-binding protein	gi	1184 11153	*T. pseudonana*	Photosynthesis	C	15518	9.6	46.8
Photosystem I p700 chlorophyll A apoprotein A	gi	1184 11096	*T. pseudonana*	Photosynthesis	C	83642	7.3	3.2
Photosystem I p700 chlorophyll A apoprotein B	gi	1184 11097	*T. pseudonana*	Photosynthesis	C	82090	7.6	5.2
Photosystem I protein F	gi	1184 11168	*T. pseudonana*	Photosynthesis	C	20362	8.9	11.9
Photosystem I protein L	gi	1184 11163	*T. pseudonana*	Photosynthesis	C	15704	9.3	18.2
Photosystem II 10 kDa phosphoprotein	gi	1184 11116	*T. pseudonana*	Photosynthesis	C	7388	6.0	21.2
Photosystem II 11 kD protein	jgi	3258	*T. pseudonana*	Photosynthesis	C	19602	9.6	5.7
Photosystem II chlorophyll A core antenna apoprotein	gi	1184 11113	*T. pseudonana*	Photosynthesis	C	56408	6.5	22.6
Photosystem II chlorophyll A core antenna apoprotein CP43	gi	1184 11149	*T. pseudonana*	Photosynthesis	C	51845	7.7	10.2
Photosystem II protein Y	gi	1184 11171	*T. pseudonana*	Photosynthesis	C	4006	12.5	22.2
Photosystem II reaction center protein D1	gi	1184 11180	*T. pseudonana*	Photosynthesis	C	39699	5.3	18.3
Photosystem II reaction center protein D2	gi	1184 11148	*T. pseudonana*	Photosynthesis	C	39064	5.6	17.7
PsbV	gi	1184 11100	*T. pseudonana*	Photosynthesis	C	17841	7.7	11.0
Putative ribosomal protein S18	jgi	26893	*T. pseudonana*	Translation	R	17159	10.8	21.9
Ribosomal protein L27	jgi	39735	*T. pseudonana*	Translation	R	16090	10.7	9.5
Ribosomal protein L5	jgi	802	*T. pseudonana*	Translation	C	35283	8.6	3.9
Ribosomal protein S12	jgi	37628	*T. pseudonana*	Translation	R	12611	6.2	9.6
Ribosomal protein S13	jgi	26221	*T. pseudonana*	Translation	R	17054	10.4	24.5
Ribosomal protein S3	jgi	28049	*T. pseudonana*	Translation	M	29161	9.3	12.2
Ribulose-1,5-bisphosphate carboxylase/ oxygenase large subunit	gi	1184 11104	*T. pseudonana*	Photosynthesis	C	54325	6.2	33.9
Ribulose-1,5-bisphosphate carboxylase/ oxygenase small subunit	gi	1184 11103	*T. pseudonana*	Photosynthesis	C	15843	5.1	19.4
Rubisco expression protein	gi	1184 11164	*T. pseudonana*	Photosynthesis	C	32381	5.9	10.5
S-adenosyl methionine	jgi	21815	*T. pseudonana*	Metabolism, one carbon	S	50359	5.2	3.2

synthetase								
S-adenosyl-L-homocysteinas protein	jgi	28496	*T. pseudonana*	Metabolism, one carbon	S	52309	5.1	11.4
SPAC22H10.12c	jgi	26136	*T. pseudonana*	GTPase	S	49514	5.4	4.9
Structural constituent of ribosome, 13 kDa	jgi	262056	*T. pseudonana*	Translation	R	13713	9.9	9.7
Structural constituent of ribosome, 14 kDa	jgi	31084	*T. pseudonana*	Translation	R	14755	10.4	9.4
Transaldolase	jgi	27187	*T. pseudonana*	Metabolism, Carbohydrate	S	34855	4.8	9.1
Translation elongation factor Tu	gi	1184 11218	*T. pseudonana*	Biosynthesis, Protein	M	44458	4.9	11.7
Tubulin beta chain	jgi	31569	*T. pseudonana*	Microtubule based movement	S	49670	4.9	2.2
Ubiquitin	jgi	40669	*T. pseudonana*	Modification, Protein	S	17567	9.9	16.3
UDP-glucose pyrophosphorylase	jgi	262059	*T. pseudonana*	Nucleotidyl-transferase activity	C	47135	5.4	1.9
Unknown	jgi	39299	*T. pseudonana*	Binding, GTP	U	20905	6.8	6.0
Unknown protein	jgi	4820	*T. pseudonana*	n.a.	S	45854	6.6	2.9
Vacuolar ATP synthase 16 kDa proteolipid subunit	jgi	2233	*T. pseudonana*	Transport, Proton	S	16720	5.6	10.8
Vacuolar ATP synthase subunit A	jgi	37123	*T. pseudonana*	Transport, Proton	S	68343	5.0	6.0
DNA-directed RNA polymerase subunit gamma	gi	3386 2040	*Prochlorococcus marinus*	Transcription	S	71995	6.6	1.7
Photosystem II PsbD protein D2	gi	3386 1713	*Prochlorococcus marinus*	Photosynthesis	S	39325	5.6	6.2

Slurry Tube Gel – GOS/Thaps Database

Protein	Annotation		Species	Function	Comp	MW	pI	SC
30S ribosomal protein S5	gi	1184 11207	*T. pseudonana*	Translation	C	19297	10.3	11.7
30S ribosomal protein S6	gi	1184 11222	*T. pseudonana*	Translation	C	12043	9.8	10.7
30S ribosomal protein S7	gi	1184 11217	*T. pseudonana*	Translation	C	17730	10.5	21.2
3-deoxy-7-phospho heptulonate synthase	jgi	2790	*T. pseudonana*	Biosynthesis, amino acid	C	53939	6.0	2.5
40S ribosomal protein S11	jgi	22535	*T. pseudonana*	Translation	R	19118	10.3	8.5
40S ribosomal protein S23	jgi	28209	*T. pseudonana*	Translation	R	15734	10.5	7.7
40S ribosomal protein SA p40	jgi	21871	*T. pseudonana*	Translation	R	27261	5.9	10.3
50S ribosomal protein L11	gi	1184 11123	*T. pseudonana*	Translation	C	14880	9.7	9.2
50S ribosomal protein L14	gi	1184 11201	*T. pseudonana*	Translation	C	13433	10.3	13.2
50S ribosomal protein L23	gi	1184 11192	*T. pseudonana*	Transferase	C	11152	9.9	10.2
50S ribosomal protein L3	gi	1184 11190	*T. pseudonana*	Translation	C	22012	10.1	26.1
50S ribosomal protein L4	gi	1184 11191	*T. pseudonana*	Translation	C	24203	10.2	7.9
60 kDa chaperonin	gi	1184	*T. pseudonana*	Metabolism	C	57361	5.2	3.2

		11188						
6-phospho gluconate dehydrogenase	jgi	33343	*T. pseudonana*	Pentose phosphate shunt	S	53348	5.6	9.8
Acetyl-CoA carboxylase	jgi	6770	*T. pseudonana*	Ligase	C	228295	5.0	1.5
Acidic ribosomal phospho protein P0	jgi	25812	*T. pseudonana*	Biosynthesis	R	34116	4.8	3.7
Adenosine triphosphatase	jgi	40156	*T. pseudonana*	Transport, Proton	S	39935	7.6	7.4
Adenylate kinase	jgi	31809	*T. pseudonana*	Metabolism, nucleic acid	S	24598	6.3	4.4
ATP synthase CF0 B chain subunit I	gi	1184 11110	*T. pseudonana*	Photosynthesis	C	20029	9.8	8.4
ATP synthase CF0 B' chain subunit II	gi	1184 11109	*T. pseudonana*	Photosynthesis	C	17373	4.6	9.6
ATP synthase CF0 C chain subunit III	gi	1184 11108	*T. pseudonana*	Photosynthesis	C	8166	5.0	50
ATP synthase CF1 alpha chain	gi	1184 11112	*T. pseudonana*	Photosynthesis	C	53989	5.0	6
ATP synthase CF1 beta chain	gi	1184 11134	*T. pseudonana*	Photosynthesis	C	51143	4.7	25.1
ATP/ADP translocator	jgi	39143	*T. pseudonana*	Transport	S	32254	9.4	9.9
ATP-dependent clp protease ATP-binding subunit	gi	1184 11220	*T. pseudonana*	Catalytic activity	C	102150	6.5	1.5
CbbX protein homolog	jgi	40193	*T. pseudonana*	Binding, ATP	C	35036	5.3	15.1
Cell division protein FtsH-like protein	gi	1184 11141	*T. pseudonana*	Photosynthesis	C	70206	5.1	3.9
CG11154-PA, isoform A	jgi	41256	*T. pseudonana*	Transport, Proton	S	53388	5.1	6.6
Chloroplast light harvesting protein isoform 12	jgi	270092	*T. pseudonana*	Photosynthesis	C	26078	5.5	6.2
Chloroplast light harvesting protein isoform 15	jgi	2845	*T. pseudonana*	Photosynthesis	C	21873	5.1	5.4
Chloroplast light harvesting protein isoform 5	jgi	32723	*T. pseudonana*	Photosynthesis	C	19175	5.2	10.6
Coatomer protein complex, subunit gamma 2	jgi	269663	*T. pseudonana*	Transport, Protein	S	100270	5.1	1.1
Cytochrome b6	gi	1184 11154	*T. pseudonana*	Photosynthesis	C	23906	9.2	6
Cytochrome f	gi	1184 11137	*T. pseudonana*	Photosynthesis	C	33988	8.2	4.1
Elongation factor 2	jgi	269148	*T. pseudonana*	GTPase	S	91887	6.0	8.7
Elongation factor alpha-like protein	jgi	41829	*T. pseudonana*	GTPase	S	49969	8.7	7.1
Enoyl-acyl carrier reductase	jgi	32860	*T. pseudonana*	Oxidation reduction	S	32813	5.1	3.5
Eukaryotic translation initiation factor 4A2 isoform 2	jgi	9716	*T. pseudonana*	Binding, DNA	S	42405	5.6	3.5
Fructose-1,6-bisphosphate aldolase precursor	jgi	428	*T. pseudonana*	Glycolysis	S	39810	4.8	8.1
Fucoxanthin chlorophyl a/c protein, 21.8 kDa	jgi	38667	*T. pseudonana*	Photosynthesis	C	21807	4.8	17.6
Fucoxanthin chlorophyll a /c protein, 20.3 kDa	jgi	38494	*T. pseudonana*	Photosynthesis	C	20354	4.5	11.6
Fucoxanthin chlorophyll a/c binding protein, 22.2	jgi	264921	*T. pseudonana*	Photosynthesis	C	22205	4.6	8.6

kDa								
Fucoxanthin chlorophyll a/c binding protein, 22.6 kDa	jgi	268127	*T. pseudonana*	Photosynthesis	C	22628	4.8	17.1
GDP-mannose 3,5-epimerase	jgi	41548	*T. pseudonana*	Coenzyme Metabolic,	S	40707	5.1	3.3
Geranyl-geranyl reductase	jgi	10234	*T. pseudonana*	Aromatic Compound	C	47230	5.9	10.3
Glyceraldehyde-3-phosphate dehydrogenase	jgi	28334	*T. pseudonana*	Glycolysis	M	36574	6.1	4.4
Glyceraldehyde-3-phosphate dehydrogenase precursor	jgi	31383	*T. pseudonana*	Glycolysis	S	39587	5.3	5.3
Glycolaldehyde transferase	jgi	21175	*T. pseudonana*	Transport	M	71708	5.0	2.3
Heat shock protein 70	jgi	269120	*T. pseudonana*	Morphogenesis, Cell	S	71187	4.8	8.1
Hsp70-type chaperone	gi	1184 11189	*T. pseudonana*	Transcription	C	65339	4.8	2.3
HSP90-like protein	jgi	22766	*T. pseudonana*	Binding, ATP	S	80966	4.7	2
Hypothetical protein CBG08717	jgi	269322	*T. pseudonana*	Transport, Proton	S	58009	5.8	7.7
Hypothetical protein CdifQ_02003487	jgi	36462	*T. pseudonana*	Metabollic	U	21340	4.8	6.7
Hypothetical protein FG01081.1	jgi	25949	*T. pseudonana*	Translation	R	20124	9.8	7.9
Iron-sulfur cluster formation ABC transporter	gi	1184 11105	*T. pseudonana*	Photosynthesis	C	54338	6.9	2.5
Ketol-acid reductoisomerase	jgi	23228	*T. pseudonana*	Biosynthesis, amino acid	S	58240	5.1	3.2
L4/L1	jgi	22610	*T. pseudonana*	Translation	R	40991	10.3	7.7
Oxygen-evolving enhancer protein 1 precursor	jgi	34830	*T. pseudonana*	Photosynthesis	S	29136	5.2	23.6
Phosphoglycerate kinase precursor	jgi	35712	*T. pseudonana*	Glycolysis	S	42256	5.0	3.2
Phospho ribulokinase	jgi	4376	*T. pseudonana*	Biosynthesis	C	42389	4.9	3.3
Photosystem I ferredoxin-binding protein	gi	1184 11153	*T. pseudonana*	Photosynthesis	C	15518	9.6	20.1
Photosystem I p700 chlorophyll A apoprotein A	gi	1184 11096	*T. pseudonana*	Photosynthesis	C	83642	7.3	3.2
Photosystem I p700 chlorophyll A apoprotein B	gi	1184 11097	*T. pseudonana*	Photosynthesis	C	82090	7.6	4.1
Photosystem I protein F	gi	1184 11168	*T. pseudonana*	Photosynthesis	C	20362	8.9	11.9
Photosystem I protein L	gi	1184 11163	*T. pseudonana*	Photosynthesis	C	15704	9.3	18.2
Photosystem II 10 kDa phosphoprotein	gi	1184 11116	*T. pseudonana*	Photosynthesis	C	7388	6.0	21.2
Photosystem II 11 kD protein	jgi	3258	*T. pseudonana*	Photosynthesis	C	19602	9.6	5.7
Photosystem II chlorophyll A core antenna apoprotein	gi	1184 11113	*T. pseudonana*	Photosynthesis	C	56408	6.5	22.6
Photosystem II reaction center protein D1	gi	1184 11180	*T. pseudonana*	Photosynthesis	C	39699	5.3	13.1
Photosystem II reaction center protein	gi	1184 11148	*T. pseudonana*	Photosynthesis	C	39064	5.6	17.9

D2								
PsbV	gi	1184 11100	*T. pseudonana*	Photosynthesis	C	17841	7.7	11
Ribosomal protein L27	jgi	39735	*T. pseudonana*	Translation	R	16090	10.7	9.5
Ribosomal protein L5	jgi	802	*T. pseudonana*	Translation	C	35283	8.6	3.9
Ribosomal protein S12	jgi	37628	*T. pseudonana*	Translation	R	12611	6.2	9.6
Ribosomal protein S18	jgi	26893	*T. pseudonana*	Translation	R	17159	10.8	21.9
Ribosomal protein S3	jgi	28049	*T. pseudonana*	Translation	M	29161	9.3	7.8
Ribulose-1,5-bisphosphate carboxylase/ oxygenase large subunit	gi	1184 11104	*T. pseudonana*	Photosynthesis	C	54325	6.2	27.3
Ribulose-1,5-bisphosphate carboxylase/ oxygenase small subunit	gi	1184 11103	*T. pseudonana*	Photosynthesis	C	15843	5.1	19.4
Rubisco expression protein	gi	1184 11164	*T. pseudonana*	Photosynthesis	C	32381	5.9	10.5
S-adenosyl methionine synthetase	jgi	21815	*T. pseudonana*	Metabolic, One carbon	S	50359	5.2	3.2
S-adenosyl-L-homocysteinas protein	jgi	28496	*T. pseudonana*	Metabolic, One carbon	S	52309	5.1	7.9
SPAC22H10.12c	jgi	26136	*T. pseudonana*	GTPase	S	49514	5.4	4.9
Structural constituent of ribosome	jgi	31084	*T. pseudonana*	Translation	R	14755	10.4	9.4
Transaldolase	jgi	27187	*T. pseudonana*	Metabolism, carbohydrate	S	34855	4.8	9.1
Translation elongation factor Tu	gi	1184 11218	*T. pseudonana*	Biosynthesis, Protein	M	44458	4.9	8.1
Vacuolar ATP synthase 16 kDa proteolipid subunit	jgi	2233	*T. pseudonana*	Transport, Proton	S	16720	5.6	10.8
Vacuolar ATP synthase subunit A	jgi	37123	*T. pseudonana*	Transport, Proton	S	68343	5.0	4.8
50S ribosomal protein L22	gi	1184 11196	*T. pseudonana*, JCVI_PEP-1096694753977	Translation	C	12986	10.3	11.3
Putative CDC48/ATPase	jgi	267952	*T. pseudonana*, JCVI_PEP-1096692712161	Binding, ATP	S	89464	4.8	7.3
Glucose-6-phosphate dehydrogenase	jgi	34514	*T. pseudonana*, JCVI_PEP-1096675866109	Metabolism, Glucose	S	57149	7.6	2.8
Citrate synthase	jgi	11411	*T. pseudonana*, JCVI_PEP-1096675019495	Transferase	M	52269	6.2	6
Core histone	jgi	3184	*T. pseudonana*, JCVI_PEP-1096673929279	Nucleosome assembly	N	15268	11.5	25.2
Photosystem II chlorophyll A core antenna apoprotein CP43	gi	1184 11149	*T. pseudonana*, JCVI_PEP-1096672603429	Photosynthesis	C	51845	7.7	16.1
GDP-mannose 4,6-dehydratase	jgi	40586	*T. pseudonana*, JCVI_PEP-1096671414083	Metabolic, GDP-mannose	S	40656	5.9	6.2
Unknown	JCVI	109668 3416369	*Alpha proteobacterium*	n.a.	S	24195	6.9	14.3
Isocitrate dehydrogenase, NADP-dependent	JCVI	109668 9319217	*Azotobacter vinelandii*	Metabolism, Isocitrate process	S	26694	8.6	7.7
2-alkenal reductase	JCVI	109666	*Bursaphe-*	Modification,	S	50833	n.a.	12.1

		6945533	*lenchus xylophilus*	RNA				
DNA-directed RNA polymerase, beta' subunit	JCVI	109666 5662829	*Candidatus Pelagibacter ubique*	Transcription	S	154508	n.a.	2
FMN-dependent dehydrogenase	JCVI	109667 4553715	*Candidatus Pelagibacter ubique*	Oxidation reduction	S	24460	8.8	9.6
Penicillin-binding protein, 1A family	JCVI	109668 7481801	*Candidatus Pelagibacter ubique*	Biosynthesis, peptidoglycan	S	31421	9.3	6.1
Adenylylsulfate reductase, alpha subunit	JCVI	109667 2730853	*Desulfovibrio vulgaris*	Oxidation reduction	S	15446	6.2	5.5
ATP synthase F1, alpha subunit	JCVI	109667 5315161	*Emiliania huxleyi*	Synthesis, ATP	C	53427	n.a.	4.2
Unknown	JCVI	109666 5982711	*Flavobacteriales bacterium ALC-1*	n.a.	S	n.a.	n.a.	8.8
Beta-tubulin	JCVI	109666 8084639	*Gastrostyla steinii*	Microtubule based movement	S	25577	4.7	12.5
Unknown	JCVI	109667 3080787	*Hydrogeno-baculum sp.*	n.a.	S	30796	9.1	4.9
Extracellular solute-binding protein, family 1	JCVI	109666 6167981	*Labrenzia aggregata*	Transport	S	70965	n.a.	4.8
Glyceraldehyde-3-phosphate dehydrogenase, type I	JCVI	109668 2420143	*Microscilla marina*	Glycolysis	S	27653	4.8	6.2
Arginino succinate lyase	JCVI	109668 1073423	*Natronomonas pharaonis*	Biosynthesis, arginine	S	22803	10.0	11.2
Unknown	JCVI	109669 4739943	*Porphyra yezoensis*	n.a.	U	44881	6.4	15.8
Actin	JCVI	109667 3373927	*Prymnesium parvum*	Binding, Protein	S	28829	n.a.	5
Fructose-bisphosphate aldolase class-I	JCVI	109667 8563861	*Pseudo-alteromonas atlantica*	Glycolysis	S	29476	4.4	4.9
Peptidase M24	JCVI	109668 2068811	*Pseudo-alteromonas atlantica*	Proteolysis	S	49553	6.3	2.1
Unknown	JCVI	109669 1696875	*Synecho-coccus sp.*	n.a.	S	32778	5.5	7.8
Actin	JCVI	109666 8934071	*Vannella ebro*	Binding, Protein	S	25450	5.4	17.5

Slurry Tube Gel – NCBI-NR Database

Protein	Annotation		Species	Function	Comp	MW	pI	SC
50S ribosomal protein L3	gi	118411190	*T. pseudonana*	Translation	C	22012	10.1	26.1
6-phospho gluconate dehydrogenase	gi	224000295	*T. pseudonana*	Pentose phosphate shunt	S	53314	5.5	3.5
Fucoxanthin chlorophyll a/c protein 2	gi	224012180	*T. pseudonana*	Photosynthesis	C	21129	4.5	11.6
Glyceraldehyde-3-phosphate dehydrogenase precursor	gi	223993043	*T. pseudonana*	Glycolysis	S	39587	5.3	5.3
Heat shock protein 70	gi	224003673	*T. pseudonana*	Heat Shock	S	71142	4.8	4.6
Oxygen-evolving enhancer protein 1 precursor	gi	224003107	*T. pseudonana*	Photosynthesis	S	29136	5.2	18.7
Photosystem II	gi	118411113	*T. pseudonana*	Photosynthesis	C	56408	6.5	14.9

chlorophyll A core antenna apoprotein								
Predicted protein	gi	224000661	*T. pseudonana*	n.a.	O	42732	4.9	7.5
Predicted protein	gi	223998096	*T. pseudonana*	n.a.	O	21907	5.8	9.2
Predicted protein	gi	224005154	*T. pseudonana*	n.a.	O	46358	4.9	2.9
Ribulose-1,5-bisphosphate carboxylase/ oxygenase large subunit	gi	118411104	*T. pseudonana*	Photosynthesis	C	54325	6.2	23.1
RL4e, ribosomal protein 4e 60S large ribosomal subunit	gi	224000902	*T. pseudonana*	Translation	R	40965	10.3	7.7
RS18, ribosomal protein 18 40S small ribosomal subunit	gi	223994887	*T. pseudonana*	Translation	R	17148	10.7	8.2
Transketolase	gi	223995033	*T. pseudonana*	Transferase	S	71662	5.0	2.3
Translation elongation factor alpha	gi	224007705	*T. pseudonana*	Biosynthesis	S	49937	8.6	4.4
Translation factor tu domain 2	gi	224002995	*T. pseudonana*	Biosynthesis, Protein	M	91826	5.9	2.2
30S ribosomal protein S7	gi	118411217	*T. pseudonana* + Dia	Translation	C	17730	10.5	10.3
60 kDa chaperonin	gi	118411188	*T. pseudonana* + Dia	Metabolism	C	57325	5.1	3.2
Acetyl-coa carboxylase	gi	224004864	*T. pseudonana* + Dia	Ligase	C	228150	5.0	0.9
Adenine nucleotide translocator; ATP/ADP translocase	gi	223993143	*T. pseudonana* + Dia	Transport, Membrane	M	32233	9.3	9.9
ATP synthase CF0 B' chain subunit II	gi	118411109	*T. pseudonana* + Dia	Photosynthesis	C	17373	4.6	9.0
ATP synthase CF1 alpha chain	gi	118411112	*T. pseudonana* + Dia	Photosynthesis	C	53989	5.0	6.0
Gdp-d-mannose 4,6-dehydratase	gi	224001660	*T. pseudonana* + Dia	Coenzyme	S	40386	5.8	4.2
Hypothetical protein THAPSDRAFT_ 10234	gi	224010635	*T. pseudonana* + Dia	n.a.	O	47200	5.8	7.3
Hypothetical protein THAPSDRAFT_ 28496	gi	224002559	*T. pseudonana* + Dia	n.a.	O	52275	5.1	5.6
Photosystem I p700 chlorophyll A apoprotein B	gi	118411097	*T. pseudonana* + Dia	Photosynthesis	C	82037	7.3	2.5
Photosystem II chlorophyll A core antenna apoprotein CP43	gi	193735617	*T. pseudonana* + Dia	Photosynthesis	C	50231	6.5	6.5
Photosystem II reaction center protein D2	gi	118411148	*T. pseudonana* + Dia	Photosynthesis	C	39064	5.6	11.7
Predicted protein	gi	223999667	*T. pseudonana* + Dia	n.a.	O	29142	9.2	5.1
Predicted protein	gi	223999031	*T. pseudonana* + Dia	n.a.	O	45984	5.6	4.7
Rubisco expression protein	gi	118411164	*T. pseudonana* + Dia	Photosynthesis	C	32381	5.9	4.8
Vacuolar proton pump alpha subunit	gi	224008993	*T. pseudonana* + Dia	Transport	S	68298	5.0	2.9
ABC transporter related protein	gi	296130848	*Cellulomonas flavigena*	Transport, Membrane	S	27730	6.8	15.1
ATP synthase CF0 C chain subunit III	gi	118411108	*Heterosigma akashiwo*	Photosynthesis	C	8166	5.0	50.0
Cytochrome f	gi	118411137	*Odontella sinensis*	Photosynthesis	C	33988	8.2	11.5

Plastidic enolase [Phaeodactylum tricornutum CCAP 1055/1]	gi	219130730	*Phaeodactylum tricornutum*	Enolase	S	51061	4.8	2.9
Predicted protein	gi	219119939	*Phaeodactylum tricornutum*	n.a.	O	21196	10.5	8.2
Predicted protein	gi	219126467	*Phaeodactylum tricornutum*	n.a.	O	26530	10.0	5.4
Protein fucoxanthin chlorophyll a/c protein	gi	219112233	*Phaeodactylum tricornutum*	Photosynthesis	C	21315	4.8	8.1
Sugar ABC transporter, periplasmic sugar-binding protein, putative	gi	110681242	*Roseobacter denitrificans* + Bac	Transport	S	64449	4.6	1.6
ATP synthase CF1 subunit beta	gi	315320486	*Thalassiosira oceanica*	Photosynthesis	C	51173	4.7	21.9
Photosystem I reaction center subunit II	gi	315320504	*Thalassiosira oceanica*	Photosynthesis	C	15324	9.4	10.8
Photosystem I reaction center subunit III	gi	315320469	*Thalassiosira oceanica*	Photosynthesis	C	20438	8.9	20.0
Ribulose-1,5-bisphosphate carboxylase/ oxygenase small subunit	gi	315320528	*Thalassiosira oceanica*	Photosynthesis	C	15948	5.0	19.4

Slurry Tube Gel – NCBI-Refined Database

Protein	Annotation		Species	Function	Comp	MW	pI	SC
30S ribosomal protein S11	gi	118411211	*T. pseudonana*	Translation	C	13812	11.0	8.5
50S ribosomal protein L22	gi	118411196	*T. pseudonana*	Translation	C	12986	10.3	11.3
50S ribosomal protein L3	gi	118411190	*T. pseudonana*	Translation	C	22012	10.1	26.1
50S ribosomal protein L4	gi	118411191	*T. pseudonana*	Translation	C	24203	10.2	7.9
6-phospho gluconate dehydrogenase	jgi	33343	*T. pseudonana*	Pentose Phosphate Shunt	S	53348	5.6	6.3
Acetyl-coa carboxylase	gi	224004864	*T. pseudonana*	Ligase	C	228150	5.0	2.0
Adenine nucleotide translocator; ATP/ADP translocase	gi	223993143	*T. pseudonana*	Transport, Membrane	M	32233	9.3	9.9
ATP synthase	gi	223998931	*T. pseudonana*	Photosynthesis	C	39910	6.7	7.4
ATP synthase CF1 delta chain	gi	118411111	*T. pseudonana*	Photosynthesis	C	21077	9.2	16.0
Fucoxanthin chlorophyll a/c protein 8	gi	223993505	*T. pseudonana*	Photosynthesis	C	22613	4.7	17.1
Heat shock protein 70	gi	224003673	*T. pseudonana*	Heat Shock	S	71142	4.8	6.1
Hypothetical protein THAPSDRAFT_ 10234	gi	224010635	*T. pseudonana*	Unknown	U	47200	5.8	7.3
Hypothetical protein THAPSDRAFT_ 28496	gi	224002559	*T. pseudonana*	Unknown	U	52275	5.1	9.8
Hypothetical protein THAPSDRAFT_	gi	224006870	*T. pseudonana*	Unknown	U	31179	4.9	10.2

7881

Mitochondrial ATPase, inner membrane	gi	224005467	*T. pseudonana*	Metabolism	M	57972	5.7	7.7
Oxygen-evolving enhancer protein 1 precursor	gi	224003107	*T. pseudonana*	Photosynthesis	S	29136	5.2	18.7
Photosystem II chlorophyll A core antenna apoprotein	gi	118411113	*T. pseudonana*	Photosynthesis	C	56408	6.5	20.6
Predicted protein	gi	223999031	*T. pseudonana*	Unknown	U	45984	5.6	11.4
Predicted protein	gi	224000942	*T. pseudonana*	Unknown	U	17555	9.8	24.2
Predicted protein	gi	223999667	*T. pseudonana*	Unknown	U	29142	9.2	7.8
Predicted protein	gi	223998096	*T. pseudonana*	Unknown	U	21907	5.8	9.2
Predicted protein	gi	224005154	*T. pseudonana*	Unknown	U	46358	4.9	2.9
Predicted protein	gi	223997122	*T. pseudonana*	Unknown	U	19590	9.5	5.7
Predicted protein	gi	223999673	*T. pseudonana*	Unknown	U	45824	6.4	2.9
Ribosomal protein 11 40S small ribosomal subunit	gi	224000754	*T. pseudonana*	Translation	R	19105	10.2	8.5
Ribosomal protein 11A 60S large ribosomal subunit	gi	224000193	*T. pseudonana*	Translation	R	20111	9.7	7.9
Ribosomal protein 18 40S small ribosomal subunit	gi	223994887	*T. pseudonana*	Translation	R	17148	10.7	14.4
Ribosomal protein 27A 60S large ribosomal subunit	gi	223995803	*T. pseudonana*	Translation	R	16080	10.6	9.5
Ribosomal protein 4e 60S large ribosomal subunit	gi	224000902	*T. pseudonana*	Translation	R	40965	10.3	7.7
Ribosomal protein 5, 60S large ribosomal subunit	gi	224011241	*T. pseudonana*	Translation	R	35261	8.5	3.9
Ribulose-1,5-bisphosphate carboxylase/ oxygenase large subunit	gi	118411104	*T. pseudonana*	Photosynthesis	C	54325	6.2	29.2
S-adenosyl methionine synthetase	gi	223998420	*T. pseudonana*	Metabolism, One-Carbon	S	50327	5.2	3.2
Transaldolase	jgi	27187	*T. pseudonana*	Metabolic, carbohydrate	S	34855	4.8	9.1
Transketolase	gi	223995033	*T. pseudonana*	Transferase	S	71662	5.0	2.3
Translation elongation factor alpha	gi	224007705	*T. pseudonana*	Biosynthesis	S	49937	8.6	7.1
Translation factor tu domain 2	gi	224002995	*T. pseudonana*	Hydrolase	S	91826	5.9	6.6
Tubulin beta	gi	223993357	*T. pseudonana*	Microtubule based movement	S	49637	4.8	2.2
30S ribosomal protein S5	gi	118411207	*T. pseudonana* + Dia	Translation	C	19297	10.3	11.9
30S ribosomal protein S7	gi	118411217	*T. pseudonana* + Dia	Translation	C	17730	10.5	21.8
50S ribosomal protein L11	gi	118411123	*T. pseudonana* + Dia	Translation	C	14880	9.7	9.2
50S ribosomal protein L14	gi	118411201	*T. pseudonana* + Dia	Translation	C	13433	10.3	13.2
60 kDa chaperonin	gi	118411188	*T. pseudonana* + Dia	Metabolism	C	57325	5.1	3.2
Actin-like protein	gi	224012529	*T. pseudonana* + Dia	Structure	S	41764	5.0	4.8

ATP synthase CF0 B chain subunit I	gi	118411110	*T. pseudonana* + Dia	Photosynthesis	C	20029	9.8	8.4
ATP synthase CF0 B' chain subunit II	gi	118411109	*T. pseudonana* + Dia	Photosynthesis	C	17373	4.6	9.6
ATP synthase CF0 C chain subunit III	gi	118411108	*T. pseudonana* + Dia	Photosynthesis	C	8166	5.0	50.0
Enoyl-reductase [NADH]	gi	223997126	*T. pseudonana* + Dia	Biosynthesis, Fatty Acid	S	32793	5.0	3.5
Eukaryotic translation initiation factor 4A	gi	224009464	*T. pseudonana* + Dia	Biosynthesis, Protein	S	42377	5.5	9.2
Heat shock protein Hsp90	gi	224002106	*T. pseudonana* + Dia	Heat shock	S	80242	4.7	2.0
Hypothetical protein THAPSDRAFT_ 267952	gi	223993867	*T. pseudonana* + Dia	Unknown	U	89406	4.8	3.1
Metalloprotease	gi	223995685	*T. pseudonana* + Dia	Proteolysis	S	61916	5.1	1.9
Photosystem I p700 chlorophyll A apoprotein A	gi	118411096	*T. pseudonana* + Dia	Photosynthesis	C	83642	7.3	3.2
Photosystem I p700 chlorophyll A apoprotein B	gi	118411097	*T. pseudonana* + Dia	Photosynthesis	C	82037	7.3	5.2
Photosystem II chlorophyll A core antenna apoprotein CP43	gi	193735617	*T. pseudonana* + Dia	Photosynthesis	C	50231	6.5	10.5
Photosystem II reaction center protein D1	gi	118411180	*T. pseudonana* + Dia	Photosynthesis	C	39699	5.3	6.5
PsbV	gi	118411100	*T. pseudonana* + Dia	Photosynthesis	C	17841	7.7	11.0
Translation elongation factor Tu	gi	118411218	*T. pseudonana* + Dia	Biosynthesis, Protein	M	44458	4.9	8.1
UDP-glucose 4-epimerase	gi	224005665	*T. pseudonana* + Dia	Metabolism, Isomerase	S	40680	5.1	3.3
Vacuolar proton pump alpha subunit	gi	224008993	*T. pseudonana* + Dia	Transport	S	68298	5.0	4.8
V-type ATPase	gi	223995041	*T. pseudonana* + Dia	Transport, Proton	S	16709	5.0	10.8
Gdp-d-mannose 4,6-dehydratase	gi	224001660	*T. pseudonana* + Dia + Bac	Coenzyme	S	40386	5.8	6.2
Hypothetical protein THAPSDRAFT_ 1456	gi	223993109	*T. pseudonana* + Dia + Bac	Unknown	U	73761	5.7	2.4
Photosystem II D2 protein	gi	126656960	*Cyanothece sp.*	Photosynthesis	C	39328	5.6	9.1
DNA-directed RNA polymerase subunit beta'	gi	108759710	*Myxococcus xanthus* + Bac	Transcription	S	156216	7.5	1.6
Cytochrome f	gi	118411137	*Odontella sinensis*	Photosynthesis	C	33988	8.2	11.5
Fucoxanthin chlorophyll a/c protein	gi	219112233	*Phaeodactylum tricornutum*	Photosynthesis	C	21315	4.8	8.1
Plastidic enolase	gi	219130730	*Phaeodactylum tricornutum*	Enolase	S	51061	4.8	2.9
Predicted protein	gi	219126467	*Phaeodactylum tricornutum*	Unknown	U	26530	10.0	5.4
Predicted protein	gi	219124224	*Phaeodactylum tricornutum*	Unknown	U	46624	4.9	2.9
Predicted protein	gi	219129995	*Phaeodactylum tricornutum*	Unknown	U	21321	10.0	11.7
Predicted protein	gi	219122924	*Phaeodactylum tricornutum*	Unknown	U	15466	11.3	9.6
Fructose-bisphosphate aldolase	gi	219114000	*Phaeodactylum tricornutum* + Dia	Hydrolase	S	43552	4.9	2.8
Rubisco expression	gi	118410970	*Phaeodactylum*	Photosynthesis	C	32728	7.8	10.7

protein			*tricornutum* + Dia					
Sugar ABC transporter	gi	110681242	*Roseobacter denitrificans* + Bac	Transport	S	64449	4.6	8.3
30S ribosomal protein S6	gi	315320561	*Thalassiosira oceanica*	Translation	C	12013	9.5	23.3
50S ribosomal protein L1	gi	315320476	*Thalassiosira oceanica*	Translation	C	25115	8.5	6.1
50S ribosomal protein L21	gi	315320545	*Thalassiosira oceanica*	Translation	C	12519	10.1	21.0
50S ribosomal protein L22	gi	315320570	*Thalassiosira oceanica*	Translation	C	13018	10.2	11.3
50S ribosomal protein L23/L25	gi	315320566	*Thalassiosira oceanica*	Translation	C	10533	9.9	15.2
ATP synthase CF1 subunit alpha	gi	315320537	*Thalassiosira oceanica*	Photosynthesis	C	53989	5.0	10.7
ATP synthase CF1 subunit beta	gi	315320486	*Thalassiosira oceanica*	Photosynthesis	C	51173	4.7	31.0
Photosystem I reaction center subunit II	gi	315320504	*Thalassiosira oceanica*	Photosynthesis	C	15324	9.4	27.3
Photosystem I reaction center subunit III	gi	315320469	*Thalassiosira oceanica*	Photosynthesis	C	20438	8.9	20.0
Ribulose-1,5-bisphosphate carboxylase/ oxygenase small subunit	gi	315320528	*Thalassiosira oceanica*	Photosynthesis	C	15948	5.0	28.8

Traditional Tube Gel – Thaps Database

Protein	Annotation		Species	Function	Comp	MW	pI	SC
-: -	jgi	10417	*T. pseudonana*	n.a.	S	22939	4.9	6.4
14-3-3	jgi	26146	*T. pseudonana*	Binding Protein domain specific	S	27869	4.6	8.9
6-phospho gluconate dehydrogenase	jgi	33343	*T. pseudonana*	Decar-boxylation	S	53348	5.6	3.5
Acetyl-CoA carboxylase	jgi	6770	*T. pseudonana*	Ligase	S	228295	5.0	1.3
Actin A	jgi	25772	*T. pseudonana*	Binding Protein	S	41791	5.0	6.4
Adenosine triphosphatase	jgi	40156	*T. pseudonana*	Transport, proton	S	39935	7.6	7.4
Amino transferase AGD2	jgi	31394	*T. pseudonana*	Biosynthesis	M	44264	4.8	2.9
ATP synthase CF0 B chain subunit I	gi	1184 11110	*T. pseudonana*	Photosynthesis	C	20029	9.8	7.8
ATP synthase CF0 B' chain subunit II	gi	1184 11109	*T. pseudonana*	Photosynthesis	C	17373	4.6	9.0
ATP synthase CF0 C chain subunit III	gi	1184 11108	*T. pseudonana*	Photosynthesis	C	8166	5.0	50.0
ATP synthase CF1 alpha chain	gi	1184 11112	*T. pseudonana*	Photosynthesis	C	53989	5.0	4.6
ATP synthase CF1 beta chain	gi	1184 11134	*T. pseudonana*	Photosynthesis	C	51143	4.7	22.6
BiP	jgi	27656	*T. pseudonana*	Morpho genesis, Cell	C	70451	4.7	4.5
Cell division protein FtsH-like protein	gi	1184 11141	*T. pseudonana*	Proteolysis	C	70206	5.1	2.0
CG11154-PA, isoform A	jgi	41256	*T. pseudonana*	Transport, Proton	S	53388	5.1	11.8
Cytochrome b6	gi	1184 11154	*T. pseudonana*	Photosynthesis	C	23906	9.2	6.0
Eukaryotic	jgi	9716	*T. pseudonana*	Binding, DNA	S	42405	5.6	3.5

translation initiation factor 4A2 isoform 2								
Fructose-1,6-bisphosphate aldolase precursor	jgi	428	*T. pseudonana*	Glycolysis	S	39810	4.8	3.0
Fucoxanthin chlorophyl a/c protein, 21.8 kDa	jgi	38667	*T. pseudonana*	Photosynthesis	C	21807	4.8	17.6
Fucoxanthin chlorophyll a /c protein, 20.3 kDa	jgi	38494	*T. pseudonana*	Photosynthesis	C	20354	4.5	31.6
Fucoxanthin chlorophyll a/c binding protein, 22.2 kDa	jgi	264921	*T. pseudonana*	Photosynthesis	C	22205	4.6	11.5
Fucoxanthin chlorophyll a/c binding protein, 22.6 kDa	jgi	268127	*T. pseudonana*	Photosynthesis	C	22628	4.8	28.6
GDP-mannose dehydratase	jgi	40586	*T. pseudonana*	Coenzyme binding	S	40412	5.9	4.2
Glyceraldehyde-3-phosphate dehydrogenase precursor	jgi	31383	*T. pseudonana*	Glycolysis	S	39587	5.3	7.2
Glycolaldehyde transferase	jgi	21175	*T. pseudonana*	Transport	M	71708	5.0	2.3
Heat shock protein 70	jgi	269120	*T. pseudonana*	Heat shock	S	71187	4.8	4.1
Heat shock protein Hsp90	jgi	6285	*T. pseudonana*	Heat shock	S	80242	4.7	2.0
Histone H4	jgi	3184	*T. pseudonana*	Binding, DNA	N	11384	11.5	29.1
HSP90-like protein	jgi	22766	*T. pseudonana*	Heat shock	S	80966	4.7	1.7
Hypothetical protein	jgi	8907	*T. pseudonana*	n.a.	U	17800	6.1	7.2
Hypothetical protein	jgi	27276	*T. pseudonana*	Transport	S	31423	6.9	2.7
Hypothetical Protein	jgi	23918	*T. pseudonana*	n.a.	S	31192	4.8	19.5
Hypothetical protein CBG01077	jgi	22792	*T. pseudonana*	Photosynthesis	S	22486	6.7	10.9
Hypothetical protein CBG08717	jgi	269322	*T. pseudonana*	Transport, Proton	S	58009	5.8	2.8
Hypothetical protein CdifQ_ 02003487	jgi	36462	*T. pseudonana*	Metabolic	U	21340	4.8	6.7
Hypothetical protein HoreDRAFT_ 1914	jgi	22483	*T. pseudonana*	n.a.	M	152913	4.9	1.5
Inorganic diphosphatase	jgi	269348	*T. pseudonana*	Phosphate metabolic process	S	29982	4.8	5.9
Isocitrate/ isopropylmalate dehydrogenase	jgi	5293	*T. pseudonana*	Oxido reductase	S	40667	4.6	2.9
Manganese superoxide dismutase	jgi	32874	*T. pseudonana*	Superoxid dismutase	M	27061	5.5	5.3
Oxygen-evolving enhancer protein 1 precursor	jgi	34830	*T. pseudonana*	Photosynthesis	S	29136	5.2	7.6
Phospho adenosine-phospho sulphate reductase	jgi	24887	*T. pseudonana*	Ox-Redox Homeostasis	C	49035	5.0	3.9
Photosystem I iron-sulfur center	gi	1184 11182	*T. pseudonana*	Photosynthesis	C	8798	6.5	16.0
Photosystem I p700 chlorophyll A apoprotein A	gi	1184 11096	*T. pseudonana*	Photosynthesis	C	83642	7.3	3.5
Photosystem II 10 kDa phosphoprotein	gi	1184 11116	*T. pseudonana*	Photosynthesis	C	7388	6.0	21.2

Photosystem II chlorophyll A core antenna apoprotein	gi	1184 11113	*T. pseudonana*	Photosynthesis	C	56408	6.5	10.2
Photosystem II chlorophyll A core antenna apoprotein CP43	gi	1184 11149	*T. pseudonana*	Photosynthesis	C	51845	7.7	6.4
Photosystem II protein Y	gi	1184 11171	*T. pseudonana*	Photosynthesis	C	4006	12.5	22.2
Photosystem II reaction center protein D1	gi	1184 11180	*T. pseudonana*	Photosynthesis	C	39699	5.3	6.4
Photosystem II reaction center protein D2	gi	1184 11148	*T. pseudonana*	Photosynthesis	C	39064	5.6	14.2
Protein phosphatase type 1	jgi	39936	*T. pseudonana*	Hydrolase	S	36502	5.1	2.8
PsbV	gi	1184 11100	*T. pseudonana*	Photosynthesis	C	17841	7.7	11.0
Putative S-adenosyl-L-homocysteinas protein	jgi	28496	*T. pseudonana*	Adenosyl-homo cysteinase activity	C	52309	5.1	3.3
Ribosomal protein S12	jgi	37628	*T. pseudonana*	Translation	C	12611	6.2	20.9
Ribulose-1,5-bisphosphate carboxylase/ oxygenase large subunit	gi	1184 11104	*T. pseudonana*	Photosynthesis	C	54325	6.2	23.1
Ribulose-1,5-bisphosphate carboxylase/ oxygenase small subunit	gi	1184 11103	*T. pseudonana*	Photosynthesis	C	15843	5.1	34.5
Transaldolase	jgi	27187	*T. pseudonana*	Metabolism, carbohydrate	S	34855	4.8	6.0
Ubiquitin	jgi	40669	*T. pseudonana*	Modification	S	17567	9.9	10.5
Unknown	jgi	39299	*T. pseudonana*	Binding, GTP	U	20905	6.8	8.2
Vacuolar ATP synthase 16 kDa proteolipid subunit	jgi	2233	*T. pseudonana*	Transport, Proton	S	16720	5.6	10.8
Vacuolar proton-inorganic pyro phosphatase	jgi	39520	*T. pseudonana*	Transport, Proton	S	70120	5.0	2.4

Traditional Tube Gel – GOS/Thaps Database

Protein	Annotation		Species	Function	Comp	MW	pI	SC
-: -	jgi	10417	*T. pseudonana*	n.a.	S	22939	4.9	6.4
14-3-3	jgi	26146	*T. pseudonana*	Binding, Protein domain specific	U	27869	4.6	4.9
6-phospho gluconate dehydrogenase	jgi	33343	*T. pseudonana*	Pentose-Phosphate Shunt	S	53348	5.6	3.5
Acetyl-CoA carboxylase	jgi	6770	*T. pseudonana*	Ligase	C	228295	5.0	0.9
Actin A	jgi	25772	*T. pseudonana*	Binding, Protein	S	41791	5.0	6.4
Adenosine triphosphatase	jgi	40156	*T. pseudonana*	Transport, Proton	S	39935	7.6	3.8
Amino transferase AGD2	jgi	31394	*T. pseudonana*	Biosynthesis	M	44264	4.8	2.9
ATP synthase CF0 B' chain subunit II	gi	118411109	*T. pseudonana*	Hydrolase	C	17373	4.6	9

ATP synthase CF0 C chain subunit III	gi	118411108	*T. pseudonana*	Transport, Proton	C	8166	5.0	50
ATP synthase CF1 alpha chain	gi	118411112	*T. pseudonana*	Transport, Proton	C	53989	5.0	2.2
Cell division protein FtsH-like protein	gi	118411141	*T. pseudonana*	Proteolysis	C	70206	5.1	2
CG11154-PA, isoform A	jgi	41256	*T. pseudonana*	Transport, Proton	S	53388	5.1	11.8
Conserved hypothetical protein	jgi	8907	*T. pseudonana*	n.a.	U	17800	6.1	7.2
Cytochrome b6	gi	118411154	*T. pseudonana*	Oxidoreductase	C	23906	9.2	6
Eukaryotic translation initiation factor 4A2 isoform 2	jgi	9716	*T. pseudonana*	Binding, DNA	S	42405	5.6	3.5
Fructose-1,6-bisphosphate aldolase precursor	jgi	428	*T. pseudonana*	Glycolysis	S	39810	4.8	3
Fucoxanthin chlorophyl a/c protein	jgi	38667	*T. pseudonana*	Photosynthesis	C	21807	4.8	17.6
Fucoxanthin chlorophyll a /c protein	jgi	38494	*T. pseudonana*	Photosynthesis	C	20354	4.5	36.8
Fucoxanthin chlorophyll a/c binding protein	jgi	264921	*T. pseudonana*	Photosynthesis	C	22205	4.6	11.5
Fucoxanthin chlorophyll a/c binding protein	jgi	268127	*T. pseudonana*	Photosynthesis	C	22628	4.8	28.6
Glycer- aldehyde-3-phosphate dehydrogenase precursor	jgi	31383	*T. pseudonana*	Glycolysis	S	39587	5.3	5.3
Glycolaldehyde transferase	jgi	21175	*T. pseudonana*	Transferase	M	71708	5.0	2.3
Hypothetical protein CBG08717	jgi	269322	*T. pseudonana*	Transport, Proton	S	58009	5.8	2.8
Hypothetical protein HoreDRAFT_ 1914	jgi	22483	*T. pseudonana*	n.a.	M	152913	4.9	1.5
Hypothetical Protein No BLAST result	jgi	23918	*T. pseudonana*	n.a.	S	31192	4.8	12.3
Inorganic diphosphatase/ magnesium ion binding / pyro phosphatase	jgi	269348	*T. pseudonana*	Diphosphate activity	S	29982	4.8	5.9
Manganese superoxide dismutase	jgi	32874	*T. pseudonana*	Superoxide metabolism	M	27061	5.5	5.3
Phospho adenosine-phospho sulphate reductase	jgi	24887	*T. pseudonana*	Metabolism	C	49035	5.0	3.9
Photosystem I p700 chlorophyll A apoprotein A	gi	118411096	*T. pseudonana*	Photosynthesis	C	83642	7.3	1.9
Photosystem II 10 kDa phosphoprotein	gi	118411116	*T. pseudonana*	Photosynthesis	C	7388	6.0	21.2
Photosystem II chlorophyll A core antenna apoprotein	gi	118411113	*T. pseudonana*	Photosynthesis	C	56408	6.5	5.5
PsbV	gi	118411100	*T. pseudonana*	Photosynthesis	C	17841	7.7	11
Ribulose-1,5-bisphosphate carboxylase/	gi	118411104	*T. pseudonana*	Photosynthesis	C	54325	6.2	16.5

oxygenase large subunit								
Ribulose-1,5-bisphosphate carboxylase/ oxygenase small subunit	gi	118411103	*T. pseudonana*	Photosynthesis	C	15843	5.1	32.4
S-adenosyl-L-homocysteinas protein	jgi	28496	*T. pseudonana*	Metabolism	S	52309	5.1	3.3
Transaldolase	jgi	27187	*T. pseudonana*	Metabolism, Carbohydrate	S	34855	4.8	4.1
Vacuolar ATP synthase 16 kDa proteolipid subunit	jgi	2233	*T. pseudonana*	Transport, Proton	S	16720	5.6	10.8
4th Best Hit:unknown	jgi	39299	*T. pseudonana*, JCVI_PEP-1096686074901	Binding, GTP	S	20905	6.8	9.4
GDP-mannose dehydratase	jgi	40586	*T. pseudonana*, JCVI_PEP-1096671414083	Catalytic	S	40412	5.9	6.2
Heat shock protein 70	jgi	269120	*T. pseudonana*, JCVI_PEP-1096666945533	Heat shock	S	71187	4.8	6.9
Histone H4	jgi	3184	*T. pseudonana*, JCVI_PEP-1096671719135	Binding, DNA	N	11384	11.5	21.1
Oxygen- evolving enhancer protein 1 precursor	jgi	34830	*T. pseudonana*, JCVI_PEP-1096690004671	Photosynthesis	S	29136	5.2	12.3
Photosystem II chlorophyll A core antenna apoprotein CP43	gi	118411149	*T. pseudonana*, JCVI_PEP-1096672603429	Photosynthesis	C	51845	7.7	10
Photosystem II reaction center protein D1	gi	118411180	*T. pseudonana*, JCVI_PEP-1096665660015	Photosynthesis	C	39699	5.3	6.4
Photosystem II reaction center protein D2	gi	118411148	*T. pseudonana*, JCVI_PEP-1096670514615	Photosynthesis	C	39064	5.6	20.8
Rubisco expression protein	gi	118411164	*T. pseudonana*, JCVI_PEP-1096675418637	Binding, ATP	C	32381	5.9	4
Ubiquitin	jgi	40669	*T. pseudonana*, JCVI_PEP-1096672450603	Modification, Protein	S	17567	9.9	20.8
Chaperone protein HtpG	JCVI	109669 8265543	*Alpha proteobacterium*	Heat shock	S	44431	5.6	3.6
Unknown	JCVI	109666 6731141	*Candidatus Pelagibacter ubique*	n.a.	U	n.a.	n.a.	6.1
Putative LicD-family phosphor transferase	JCVI	109666 9782619	*Clostridium spiroforme*	Transferase	S	20956	8.9	7.1
Unknown	JCVI	109669 1631181	*Drosophila virilis*	n.a.	U	8258	10.4	24.7
Translation elongation factor tu	JCVI	109669 2481867	*Emiliania huxleyi*	Biosynthesis, Protein	M	24010	4.9	7.3
ATP synthase F1, alpha subunit	JCVI	109667 5315161	*Emiliania huxleyi*	Transport, Proton	C	33876	5.7	4.2
Unknown	JCVI	109668 1477117	*Labrenzia alexandrii*	n.a.	U	33488	9.4	5.7
Membrane-bound dehydrogenase domain	JCVI	109668 8183991	*Pirellula staleyi*	Signalling	S	17012	4.7	10.5
Unknown	JCVI	109668 7650571	*Plasmodium falciparum*	n.a.	U	24818	4.2	8
Ribulose bisphosphate carboxylase large	JCVI	109667 3972533	*Pleurochrysis carterae*	Photosynthesis	C	9029	5.3	22.5

chain, catalytic domain								
Actin	JCVI	109667 3373927	*Prymnesium parvum*	Binding, Protein	S	28829	n.a.	19
Metal- dependent phosphor hydrolase, puative, containing HD region	JCVI	109666 6648519	*Pyrococcus abyssi*	Hydrolase	S	28754	6.0	4
Dihydroxy- acid dehydratase	JCVI	109667 0790861	*Rhodococcus sp.*	Hydrolase	S	26423	4.8	8.3
ATPase associated with various cellular activities	JCVI	109669 5972949	*Rubrobacter xylanophilus*	Biosynthesis, Auxin	S	33480	4.6	5.1
Ribulose bisphosphate carboxylase large chain, catalytic domain	JCVI	109668 4630061	*Sphaerosorus composita*	Photosynthesis	C	20481	6.6	10.3
Photosystem q(b) protein	JCVI	109669 1696875	*Synechococcus sp.*	Photosynthesis	C	32778	5.5	7.8

Traditional Tube Gel – NCBI-NR Database

Protein	Annotation		Species	Function	Comp	MW	pI	SC
14-3-3	gi	223998024	*T. pseudonana*	n.a.	O	27852	4.6	4.9
6-phospho gluconate dehydrogenase	gi	224000295	*T. pseudonana*	Pentose phosphate shunt	S	53314	5.5	3.5
Fucoxanthin chlorophyll a/c protein 2	gi	224012180	*T. pseudonana*	Photosynthesis	C	21129	4.5	32.1
Fucoxanthin chlorophyll a/c protein 4	gi	224012385	*T. pseudonana*	Photosynthesis	C	21793	4.7	6.9
Fucoxanthin chlorophyll a/c protein 8	gi	223993505	*T. pseudonana*	Photosynthesis	C	22613	4.7	24.3
Fucoxanthin chlorophyll a/c protein, LI818 clade	gi	224013064	*T. pseudonana*	Photosynthesis	C	22177	4.5	11.5
Hypothetical protein THAPSDRAFT_ 28496	gi	224002559	*T. pseudonana*	n.a.	O	52275	5.1	3.3
Phospho adenosine-phospho sulphate reductase	gi	224009658	*T. pseudonana*	Transferase	S	49003	5.0	3.9
Photosystem II chlorophyll A core antenna apoprotein	gi	118411113	*T. pseudonana*	Photosynthesis	C	56408	6.5	5.5
Predicted protein	gi	224006446	*T. pseudonana*	n.a.	O	31173	4.7	12.3
Predicted protein	gi	224000667	*T. pseudonana*	n.a.	O	152814	4.9	1.5
Predicted protein	gi	224013965	*T. pseudonana*	n.a.	O	17789	5.8	7.2
Predicted protein	gi	224014104	*T. pseudonana*	n.a.	O	22925	4.9	6.4
Predicted protein	gi	224000942	*T. pseudonana*	n.a.	O	17555	9.8	21.0
Transketolase	gi	223995033	*T. pseudonana*	Transferase	S	71662	5.0	2.3
ATP synthase CF0 B' chain subunit II	gi	118411109	*T. pseudonana* + Dia	Photosynthesis	C	17373	4.6	9.0
ATP synthase CF0 C chain subunit III	gi	118411108	*T. pseudonana* + Dia	Photosynthesis	C	8166	5.0	50.0

Eukaryotic translation initiation factor 4A	gi	224009464	*T. pseudonana* + Dia	Biosynthesis, Protein	S	42377	5.5	3.5
Photosystem II reaction center protein D2	gi	118411148	*T. pseudonana* + Dia	Photosynthesis	C	39039	5.5	11.7
Predicted protein	gi	224000661	*T. pseudonana* + Dia	n.a.	O	42732	4.9	2.8
Ribulose-1,5-bisphosphate carboxylase/ oxygenase large subunit	gi	118411104	*T. pseudonana* + Dia	Photosynthesis	C	54325	6.2	14.7
Vacuolar ATP synthase	gi	223996057	*T. pseudonana* + Dia	Transport, Proton	S	16709	5.0	10.8
Hypothetical protein DEFDS_0950	gi	291279343	*Deferribacter desulfuricans*	n.a.	O	49169	5.8	2.9
Fucoxanthin chlorophyll a/c protein	gi	219112233	*Phaeodactylum tricornutum*	Photosynthesis	C	21315	4.8	8.1
Glyceraldehyde-3-phosphate dehydrogenase precursor	gi	219123978	*Phaeodactylum tricornutum*	Glycolysis	S	40157	5.0	3.2
Predicted protein	gi	219126955	*Phaeodactylum tricornutum*	n.a.	O	47992	4.9	4.5
Photosystem II protein D1	gi	108773067	*Scenedesmus obliquus* + Bac	Photosynthesis	C	38868	5.2	6.7
Hypothetical protein	gi	115620435	*Strongylo-centrotus purpuratus*	n.a.	O	56750	4.0	3.7
ATP synthase CF1 subunit alpha	gi	315320537	*Thalassiosira oceanica*	Photosynthesis	C	54015	4.8	6.8
ATP synthase CF1 subunit beta	gi	315320486	*Thalassiosira oceanica*	Photosynthesis	C	51173	4.7	19.0
Ribulose-1,5-bisphosphate carboxylase/ oxygenase small subunit	gi	315320528	*Thalassiosira oceanica*	Photosynthesis	C	15948	5.0	42.4

Direct Digest – Thaps Database

Protein	Annotation		Species	Function	Comp	MW	pI	SC
CG11154-PA, isoform A	jgi	41256	*T. pseudonana*	Transport, Proton	S	53388	5.1	2.4
Chloroplast ribose-5-phosphate isomerase	jgi	32332	*T. pseudonana*	Pentose-phosphate shunt	C	25606	4.8	5.3
Fucoxanthin chlorophyl a/c protein, 21.8 kDa	jgi	38667	*T. pseudonana*	Photosynthesis	C	21807	4.8	6.9
Fucoxanthin chlorophyll a/c binding protein, 22.6 kDa	jgi	268127	*T. pseudonana*	Photosynthesis	C	22628	4.8	6.2
Oxygen-evolving enhancer protein 1 precursor	jgi	34830	*T. pseudonana*	Photosynthesis	S	29136	5.2	7.6
Photosystem II reaction center protein D2	gi	118411148	*T. pseudonana*	Photosynthesis	C	39064	5.6	2.3

Direct Digest – GOS/Thaps Database

Protein	Annotation		Species	Function	Comp	MW	pI	SC
Hypothetical protein BCAM1621	JCVI	109668 0884079	*Burkholderia cenocepacia*	Unknown	S	35436	5.1	11.3
Hypothetical protein CPS_4797	JCVI	109669 5385783	*Colwellia psychrerythraea*	Unknown	S	28415	8.0	13.3
PAS domain ATPase-like protein	JCVI	109669 0628741	*Labrenzia aggregata*	Binding, ATP	S	34158	7.0	11.2
Tetratri- copeptide	JCVI	109669 6510105	*Acidovorax delafieldii*	Transporter, Membrane	S	59735	5.6	5.2

Direct Digest – NCBI-NR Database

Protein	Annotation		Species	Function	Comp	MW	pI	SC
Fucoxanthin chlorophyll a/c protein 4	gi	224012385	*T. pseudonana*	Photosynthesis	C	21793	4.7	6.9
Fucoxanthin chlorophyll a/c protein 8	gi	223993505	*T. pseudonana*	Photosynthesis	C	22613	4.7	6.2
Dihydro-dipicolinate synthetase	gi	163792790	*Alpha proteobacterium*	Lyase activity	S	32740	5.9	4.2
Reduced coenzyme oxidoreductase	gi	116672470	*Arthrobacter sp.*	Oxidoreductase	S	21309	4.6	15.2
Extracellular metallo- protease	gi	16077293	*Bacillus subtilis*	Proteolysis	S	33821	9.0	5.4
TraG domain containing protein	gi	116687191	*Burkholderia cenocepacia*	Unknown	U	136835	5.5	1.0
RNA polymerase sigma 28 subunit	gi	220909315	*Cyanothece sp.*	Transcription	S	30003	7.8	8.8
Diguanylate cyclase	gi	317052573	*Desulfur-ispirillum indicum*	Phosphorus-oxygen lyase activity	S	65559	5.7	2.8
Preprotein translocase subunit SecA	gi	120434894	*Gramella forsetii*	Transport, Membrane	S	127481	5.6	1.5
ATPase	gi	119503984	*Gamma proteobacterium*	ATP biosynthesis	S	48480	4.4	8.2
50S ribosomal protein L13P	gi	282164316	*Methanocella paludicola*	Translation	R	15244	9.8	10.1
Glutamyl-tRNA reductase	gi	18312038	*Pyrobaculum aerophilum*	Reductase	S	43603	8.1	3.0
Formate C-acetyl-transferase	gi	294678387	*Rhodobacter capsulatus*	metabolism, glucose	S	82641	5.6	2.0
Fructose-bisphosphate aldolase, class II	gi	284035188	*Spirosoma linguale*	Glycolysis	S	39499	5.2	5.6
Proton glutamate symport protein	gi	197335832	*Vibrio fischeri*	Na dicarboxylate symporter activity	S	41347	5.1	9.6
ZYRO0 A11528p	gi	254577759	*Zygo-saccharomyces rouxii*	Cell communication	S	74290	4.9	2.2

Slurry Flat Gel – Thaps Database

Protein	Annotation		Species	Function	Comp	MW	pI	SC
-: -	jgi	10417	*T. pseudonana*	Unknown	S	22939	4.9	6.4
14-3-3	jgi	26146	*T. pseudonana*	Binding, Protein	S	27869	4.6	4.5
6-phospho-gluconate dehydrogenase	jgi	33343	*T. pseudonana*	Pentose Phosphate Shunt	S	53348	5.6	2.9
Adenosine-triphosphatase	jgi	40156	*T. pseudonana*	Transport, Proton	S	39935	7.6	7.4
ATP synthase CF0 B' chain subunit II	gi	118411109	*T. pseudonana*	Photosynthesis	C	17373	4.6	9.0
ATP synthase CF0 C chain subunit III	gi	118411108	*T. pseudonana*	Photosynthesis	C	8166	5.0	26.8
ATP synthase CF1 alpha chain	gi	118411112	*T. pseudonana*	Photosynthesis	C	53989	5.0	11.5
ATP synthase CF1 beta chain	gi	118411134	*T. pseudonana*	Photosynthesis	C	51143	4.7	17.3
Cytochrome b559 alpha chain	gi	118411160	*T. pseudonana*	Photosynthesis	C	9514	5.6	11.9
Cytochrome b6	gi	118411154	*T. pseudonana*	Photosynthesis	C	23906	9.2	6.0
Ferredoxin	gi	118411098	*T. pseudonana*	Photosynthesis	C	10707	3.9	14.1
Fucoxanthin chlorophyl a/c protein	jgi	38667	*T. pseudonana*	Photosynthesis	C	21807	4.8	6.9
Fucoxanthin chlorophyll a/c binding protein	jgi	268127	*T. pseudonana*	Photosynthesis	C	22628	4.8	24.8
Glycer-aldehyde-3-phosphate dehydrogenase precursor	jgi	31383	*T. pseudonana*	Glycolysis	S	39587	5.3	8.3
Glycol-aldehyde-transferase	jgi	21175	*T. pseudonana*	Transport	M	71708	5.0	4.1
Histone H4	jgi	3184	*T. pseudonana*	Binding, DNA	N	11384	11.5	9.7
HSP90-like protein	jgi	22766	*T. pseudonana*	heat shock	S	80966	4.7	1.7
Hypothetical Protein No BLAST result	jgi	23918	*T. pseudonana*	Unknown	U	31192	4.8	16.7
Manganese superoxide dismutase	jgi	32874	*T. pseudonana*	Binding, Mn	M	27061	5.5	5.3
Phospho-glycerate kinase precursor	jgi	35712	*T. pseudonana*	Glycolysis	S	42256	5.0	4.5
Phospho-ribosyl-pyro-phosphate synthetase	jgi	26109	*T. pseudonana*	Biosynthesis, Nucleotide	S	33425	8.4	3.2
Photosystem II chlorophyll A core antenna apoprotein	gi	118411113	*T. pseudonana*	Photosynthesis	C	56408	6.5	2.0
Photosystem II chlorophyll A core antenna apoprotein CP43	gi	118411149	*T. pseudonana*	Photosynthesis	C	51845	7.7	6.4
Photosystem II protein Y	gi	118411171	*T. pseudonana*	Photosynthesis	C	4006	12.5	22.2
Photosystem II reaction center protein D2	gi	118411148	*T. pseudonana*	Photosynthesis	C	39064	5.6	11.4
PREDICTED: similar to CG11154-PA, isoform A	jgi	41256	*T. pseudonana*	Transport, Proton	S	53388	5.1	7.4
Ribulose-1,5-bisphosphate	gi	118411104	*T. pseudonana*	Photosynthesis	C	54325	6.2	15.9

carboxylase/ oxygenase large subunit								
Ribulose-1,5-bisphosphate carboxylase/ oxygenase small subunit	gi	118411103	*T. pseudonana*	Photosynthesis	C	15843	5.1	27.3
Transaldolase	jgi	27187	*T. pseudonana*	Metabolism, Carbohydrate	S	34855	4.8	4.1
Ubiquitin	jgi	40669	*T. pseudonana*	Modification, Protein	S	17567	9.9	10.5
Photosystem II reaction center protein D1	gi	118411180	*T. pseudonana, Prochlorococcus marinus*	Photosynthesis	C	39699	5.3	3.9

APPENDIX 4-1

Summary list of all proteins identified in each sample. Includes identified species, biological function, cellular compartment (Comp): C = Chloroplast; S = Secretory; M = Mitochondria; N = Nucleus; U = Uncharacterized Compartment), molecular weight (MW), isoelectric point (pI) sequence coverage, percent sequence coverage (Seq Cov).

Chl Max

Protein	Annotation		Species	Function	Comp	MW	pI	SC
26S proteasome AAA-ATPase subunit RPT3	jgi	24475	Thalassiosira pseudonana	Photosynthesis	S	42455	6.0	4.8
30S ribosomal protein S1	jgi	15259	Thalassiosira pseudonana	Translation	S	31770	4.6	7.7
30S ribosomal protein S11	gi	118411211	Thalassiosira pseudonana	Translation	C	13821	11.3	13.8
30S ribosomal protein S18	gi	118411132	Thalassiosira pseudonana	Translation	C	8155	10.8	31.9
30S ribosomal protein S3	gi	118411197	Thalassiosira pseudonana	Translation	C	24091	9.2	6.1
30S ribosomal protein S7	gi	118411217	Thalassiosira pseudonana	Translation	C	17730	10.5	21.2
30S ribosomal protein S8	gi	118411204	Thalassiosira pseudonana	Translation	C	14805	9.4	8.3
3-deoxy-7-phosphoheptulonate synthase	jgi	2790	Thalassiosira pseudonana	Biosynthesis	C	53939	6.0	13.2
3-phosphoshikimate 1-carboxyvinyltransferase	jgi	33008	Thalassiosira pseudonana	Transport	S	47333	4.6	5.6
40S ribosomal protein S17	jgi	37809	Thalassiosira pseudonana	Translation	C	14145	10.0	9.8
40S ribosomal protein S5	jgi	29955	Thalassiosira pseudonana	Translation	S	24397	6.9	6.0
40S ribosomal protein S9	jgi	268651	Thalassiosira pseudonana	Binding rRNA	S	21764	10.2	8.9
40S ribosomal protein SA p40	jgi	21871	Thalassiosira pseudonana	Translation	S	27261	5.9	18.9
50S ribosomal protein L11	gi	118411123	Thalassiosira pseudonana	Translation	C	14880	9.7	9.2
50S ribosomal protein L14	gi	118411201	Thalassiosira pseudonana	Translation	C	13433	10.3	27.3
50S ribosomal protein L16	gi	118411198	Thalassiosira pseudonana	Translation	C	15581	11.0	10.2
50S ribosomal protein L2	gi	118411193	Thalassiosira pseudonana	Translation	C	30675	10.9	6.5
50S ribosomal protein L21	gi	118411174	Thalassiosira pseudonana	Binding RNA	C	12433	10.1	8.6
50S ribosomal protein L3	gi	118411190	Thalassiosira pseudonana	Translation	C	22012	10.1	19.3
60 kDa chaperonin	gi	118411188	Thalassiosira pseudonana	Binding Protein	C	57361	5.2	15.6
60s Acidic ribosomal protein	jgi	3463	Thalassiosira pseudonana	Translation	S	27213	4.6	6.1
6-phosphogluconate dehydrogenase	jgi	33343	Thalassiosira pseudonana	Dehydrogenase	S	53348	5.6	10.8
Abnormal wing discs CG2210-PA	jgi	6290	Thalassiosira pseudonana	Biosynthesis	U	17236	5.5	30.3
Acetyl-CoA carboxylase	jgi	6770	Thalassiosira pseudonana	Metabolic Process	C	228295	5.0	2.3
Acidic ribosomal	jgi	25812	Thalassiosira	Biosynthesis	S	34116	4.8	3.7

phosphoprotein P0			pseudonana	Ribosome				
Actin A	jgi	25772	Thalassiosira pseudonana	Binding Protein	S	41791	5.0	19.9
Adenosinetriphosphatase	jgi	40156	Thalassiosira pseudonana	Transport Proton	S	39935	7.6	20.2
ALA dehydratase	jgi	5240	Thalassiosira pseudonana	Biosynthesis	S	40419	5.0	7.4
Aminotransferase AGD2	jgi	31394	Thalassiosira pseudonana	Transport	S	44264	4.8	11.5
Arginyl-tRNA synthetase	jgi	40028	Thalassiosira pseudonana	translation	S	66081	4.9	2.7
ArgJ family protein	gi	71083219	Candidatus Pelagibacter ubique	Biosynthesis	S	43114	9.5	3.4
Aromatic amino acid family biosynthesis-related protein	jgi	268552	Thalassiosira pseudonana	Biosynthesis	C	66486	5.0	3.8
ATP binding / protein binding	jgi	23102	Thalassiosira pseudonana	Binding Protein	S	59170	5.3	5.3
ATP synthase CF0 B chain subunit I	gi	118411110	Thalassiosira pseudonana	Transport Proton	C	20029	9.8	8.4
ATP synthase CF0 B' chain subunit II	gi	118411109	Thalassiosira pseudonana	Transport Proton	C	17373	4.6	29.5
ATP synthase CF1 alpha chain	gi	118411112	Thalassiosira pseudonana	Transport Ion	C	53989	5.0	35.0
ATP synthase CF1 beta chain	gi	118411134	Thalassiosira pseudonana	Transport Ion	C	51143	4.7	51.1
ATP synthase CF1 delta chain	gi	118411111	Thalassiosira pseudonana	Transport Ion	C	21077	9.2	8.0
ATP/ADP translocator	jgi	39143	Thalassiosira pseudonana	Transport	M	32254	9.4	4.0
ATPase, E1-E2 type	jgi	262679	Thalassiosira pseudonana	Transport Cation	S	99192	5.6	1.5
ATP-dependent clp protease ATP-binding subunit	gi	118411220	Thalassiosira pseudonana	Catalysis	C	102150	6.5	1.5
ATP-sulfurylase	jgi	1326	Thalassiosira pseudonana	Metabolic Process	S	45362	5.2	2.0
BiP	jgi	27656	Thalassiosira pseudonana	Cell Morphogenesis	C	70451	4.7	27.6
Catalytic	jgi	41733	Thalassiosira pseudonana	Biosynthesis	C	71455	5.4	9.3
CbbX protein homolog	jgi	40193	Thalassiosira pseudonana	Biosynthesis	C	35036	5.3	25.1
CDC48/ATPase	jgi	267952	Thalassiosira pseudonana	Binding ATP	S	89464	4.8	4.4
cell division protein FtsH2	jgi	31930	Thalassiosira pseudonana	Binding Zn	C	61956	5.1	11.9
Cell division protein FtsH-like protein	gi	118411141	Thalassiosira pseudonana	Proteolysis	C	70206	5.1	15.9
CG11154-PA, isoform A	jgi	41256	Thalassiosira pseudonana	Transport Proton	U	53388	5.1	50.8
CG17332-PA, isoform A isoform 1	jgi	263135	Thalassiosira pseudonana	Transport Proton	U	33233	5.7	5.4
Chloroplast 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase precursor	jgi	29228	Thalassiosira pseudonana	Biosynthesis	C	75736	4.9	2.6
Chloroplast clp protease P	jgi	1738	Thalassiosira pseudonana	Proteolysis	C	28223	4.8	11.1
Chloroplast coproporphyrinogen III oxidase	jgi	31012	Thalassiosira pseudonana	Biosynthesis	C	34001	5.4	5.7
Chloroplast cysteine synthase 1 precursor	jgi	31829	Thalassiosira pseudonana	Metabolic Process	C	33261	5.1	6.9
Chloroplast light harvesting protein isoform 12	jgi	270092	Thalassiosira pseudonana	Photosynthesis	C	18463	4.6	6.2
Chloroplast light	jgi	33606	Thalassiosira	Photosynthesis	C	26078	5.5	13.5

harvesting protein isoform 12			pseudonana					
Chloroplast light harvesting protein isoform 15	jgi	2845	Thalassiosira pseudonana	Photosynthesis	C	21873	5.1	5.4
Chloroplast O-acetyl-serine lyase	jgi	267987	Thalassiosira pseudonana	Metabolic Process	C	38072	5.7	4.8
Chorismate synthase	jgi	38964	Thalassiosira pseudonana	Biosynthesis	C	44051	5.5	4.9
CPN60 protein	jgi	23329	Thalassiosira pseudonana	Binding Protein	C	59177	4.7	5.6
Cytochrome b559 alpha chain	gi	118411160	Thalassiosira pseudonana	Photosynthesis	C	9514	5.6	25.0
Cytochrome c-550	gi	118411100	Thalassiosira pseudonana	Electron Transport	C	17841	7.7	11.0
Cytochrome f	gi	118411137	Thalassiosira pseudonana	Photosynthesis	C	33988	8.2	28.3
Cytosolic ribosomal protein S8	jgi	29825	Thalassiosira pseudonana	Translation	S	22604	10.5	7.5
D-3-phosphoglycerate dehydrogenase	jgi	25130	Thalassiosira pseudonana	Oxidoreductase	C	50144	6.4	4.3
Diaminopimelate decarboxylase	gi	71083118	Candidatus Pelagibacter ubique	Biosynthesis, amino acid	S	45589	9.9	4.5
DNA binding	jgi	29950	Thalassiosira pseudonana	Binding DNA	C	15300	11.3	8.1
Domain specific binding protein 14-3-3	jgi	26146	Thalassiosira pseudonana	Binding Protein	S	27869	4.6	23.1
Elongation factor 2	jgi	269148	Thalassiosira pseudonana	Translation	S	91887	6.0	4.2
Elongation factor alpha-like protein	jgi	41829	Thalassiosira pseudonana	Translation	S	49969	8.7	10.6
Endoplasmic reticulum membrane fusion protein	jgi	40348	Thalassiosira pseudonana	Transcription	S	74221	4.7	2.4
Enolase	jgi	40771	Thalassiosira pseudonana	Glycolysis	S	46744	4.9	21.1
Enolase 2	jgi	40391	Thalassiosira pseudonana	Glycolysis	S	46547	4.8	21.1
Enoyl-acyl carrier reductase	jgi	32860	Thalassiosira pseudonana	Oxidation Reduction	S	32813	5.1	25.0
ENSANGP00000020417	jgi	354	Thalassiosira pseudonana	Binding, DNA	M	41939	6.9	4.3
Eukaryotic translation initiation factor 4A2 isoform 2	jgi	9716	Thalassiosira pseudonana	Binding DNA	S	42405	5.6	19.5
F0F1 ATP synthase subunit alpha	gi	33862007	Prochlorococcus marinus	Transport Proton	S	54306	4.9	4.6
F0F1 ATP synthase subunit beta	gi	33861994	Prochlorococcus marinus	Transport Proton	S	51948	5.0	13.3
F0F1 ATP synthase subunit beta	gi	71082935	Candidatus Pelagibacter ubique	Transport Proton	S	50696	4.9	9.7
Ferredoxin component	jgi	29842	Thalassiosira pseudonana	Oxidoreductase	C	18511	8.9	6.7
Ferredoxin-dependent glutamate synthase	jgi	269900	Thalassiosira pseudonana	Metabolic Process	C	178531	5.4	5.6
FeS assembly protein SufD	jgi	268364	Thalassiosira pseudonana	Binding Protein	S	31963	5.2	5.8
Formylglycineamide ribotide amidotransferase	jgi	30301	Thalassiosira pseudonana	Catalysis	S	143746	5.0	1.4
Fructose-1,6-bisphosphate aldolase precursor	jgi	428	Thalassiosira pseudonana	Glycolysis	S	39810	4.8	19.7
Fructose-bisphosphatase	jgi	264556	Thalassiosira pseudonana	Metabolic Process Carbohydrate	C	33667	5.3	5.1
Fucoxanthin chlorophyl a/c protein	jgi	38667	Thalassiosira pseudonana	Photosynthesis	C	21807	4.8	17.6
Fucoxanthin chlorophyl a/c protein	jgi	38715	Thalassiosira pseudonana	Photosynthesis	C	20718	4.9	19.3

Fucoxanthin chlorophyll a /c protein	jgi	38494	Thalassiosira pseudonana	Photosynthesis	C	20354	4.5	17.5
Fucoxanthin chlorophyll a /c protein	jgi	42962	Thalassiosira pseudonana	Photosynthesis	C	21515	5.1	18.4
Fucoxanthin chlorophyll a/c binding protein	jgi	12097	Thalassiosira pseudonana	Photosynthesis	C	27330	4.8	8.2
Fucoxanthin chlorophyll a/c binding protein	jgi	264921	Thalassiosira pseudonana	Photosynthesis	C	22205	4.6	8.6
Fucoxanthin chlorophyll a/c binding protein	jgi	268127	Thalassiosira pseudonana	Photosynthesis	C	22628	4.8	17.1
Fucoxanthin-chlorophyll a/c light-harvesting protein	jgi	33018	Thalassiosira pseudonana	Photosynthesis	C	21786	5.4	4.0
G protein beta subunit	jgi	26063	Thalassiosira pseudonana	rRNA Processing	S	35960	6.2	23.0
GDP-mannose dehydratase	jgi	40586	Thalassiosira pseudonana	Catalysis	S	40412	5.9	4.2
Geranyl-geranyl reductase	jgi	10234	Thalassiosira pseudonana	Electron Transport	C	47230	5.9	20.0
Glucose-6-phosphate isomerase	jgi	38266	Thalassiosira pseudonana	Glycolysis	S	61689	5.9	2.7
Glutamate 1-semialdehyde 2,1-aminomutase	jgi	575	Thalassiosira pseudonana	Binding Phosphate	S	43658	5.5	7.1
Glutamine synthetase	jgi	26051	Thalassiosira pseudonana	Biosynthesis	C	45620	5.2	5.8
Glyceraldehyde-3-phosphate dehydrogenase	jgi	28334	Thalassiosira pseudonana	Glycolysis	C	36574	6.1	8.4
Glyceraldehyde-3-phosphate dehydrogenase precursor	jgi	31383	Thalassiosira pseudonana	Glycolysis	S	39587	5.3	29.6
Glycolaldehydetransferase	jgi	21175	Thalassiosira pseudonana	Transport	M	71708	5.0	25.6
Heat shock protein 60	jgi	38191	Thalassiosira pseudonana	Folding Protein	S	58525	4.8	6.8
Heat shock protein 70	jgi	269120	Thalassiosira pseudonana	Folding Protein	S	71187	4.8	10.1
Heat shock protein 83	jgi	268500	Thalassiosira pseudonana	Folding Protein	S	86014	4.6	1.8
Heat shock protein Hsp90	jgi	6285	Thalassiosira pseudonana	Folding Protein	S	80242	4.7	7.5
Histone H2A.1	jgi	19793	Thalassiosira pseudonana	Binding DNA	N	13053	10.4	7.3
Histone H4	jgi	3184	Thalassiosira pseudonana	Binding DNA	N	11384	11.5	41.7
Hsp70-type chaperone	gi	118411189	Thalassiosira pseudonana	Transcription	C	65339	4.8	7.0
HSP90-like protein	jgi	22766	Thalassiosira pseudonana	Folding Protein	S	80966	4.7	2.0
Hypothetical protein	jgi	26224	Thalassiosira pseudonana	n.a.	S	31116	8.9	7.7
Hypothetical protein	jgi	38221	Thalassiosira pseudonana	n.a.	U	14645	5.6	9.3
Hypothetical Protein	jgi	23918	Thalassiosira pseudonana	n.a.	C	31192	4.8	11.3
Hypothetical Protein	jgi	24512	Thalassiosira pseudonana	n.a.	C	41414	6.0	4.8
Hypothetical Protein	jgi	6441	Thalassiosira pseudonana	n.a.	S	26822	5.4	49.8
Hypothetical protein AN1993.2	jgi	31424	Thalassiosira pseudonana	Transport	S	45198	6.2	3.8
Hypothetical protein CBG01077	jgi	22792	Thalassiosira pseudonana	Transport	S	22486	6.7	13.9
Hypothetical protein CBG08717	jgi	269322	Thalassiosira pseudonana	Transport Proton	S	58009	5.8	19.4
Hypothetical protein DDB0218359	jgi	24710	Thalassiosira pseudonana	n.a.	C	46717	4.6	3.8
Hypothetical protein DEHA0F19712g	jgi	27352	Thalassiosira pseudonana	Binding DNA	N	13064	10.3	16.7

Hypothetical protein FG01081.1	jgi	25949	Thalassiosira pseudonana	Translation	U	20124	9.8	7.9
Hypothetical protein LOC496448	jgi	269540	Thalassiosira pseudonana	Binding Protein	S	192707	5.5	0.9
Importin alpha 1	jgi	43097	Thalassiosira pseudonana	Transport Protein	S	60533	4.9	3.2
Inorganic diphosphatase/ magnesium ion binding / pyrophosphatase	jgi	269348	Thalassiosira pseudonana	Metabolic Process Phosphate	S	29982	4.8	5.5
Integrin beta 4 binding protein	jgi	29782	Thalassiosira pseudonana	Binding Ribosome	S	27279	5.0	9.4
Isocitrate/isopropylmalate dehydrogenase	jgi	5293	Thalassiosira pseudonana	Oxidoreductase	S	40667	4.6	12.5
Ketol-acid reductoisomerase	jgi	23228	Thalassiosira pseudonana	Oxidation Reduction	C	58240	5.1	8.6
L4/L1	jgi	22610	Thalassiosira pseudonana	Translation	S	40991	10.3	7.7
Magnesium-chelatase subunit I	gi	118411138	Thalassiosira pseudonana	Photosynthesis	C	39500	5.0	6.5
Malate dehydrogenase	jgi	20726	Thalassiosira pseudonana	Oxidoreductase	S	36724	6.3	7.4
Manganese superoxide dismutase	jgi	32874	Thalassiosira pseudonana	Metabolic Process	M	27061	5.5	14.8
Molecular chaperone DnaK	jgi	269240	Thalassiosira pseudonana	Morphogenesis, Cell	S	72207	5.0	2.4
Molecular chaperone DnaK2, heat shock protein hsp70-2	gi	33862260	Prochlorococcus marinus	Folding Protein	S	68202	4.8	2.7
Myo-inositol dehydrogenase precursor	jgi	1049	Thalassiosira pseudonana	Biosynthesis	S	48134	5.0	6.2
Nucleoside diphosphate kinase	jgi	31091	Thalassiosira pseudonana	Biosynthesis	C	16597	5.8	17.2
Nucleoside diphosphate kinase	jgi	12070	Thalassiosira pseudonana	Biosynthesis	C	16917	8.3	18.8
Oxygen-evolving enhancer protein 1 precursor	jgi	34830	Thalassiosira pseudonana	Photosynthesis	C	29136	5.2	27.3
PAaA	jgi	24864	Thalassiosira pseudonana	Transport Proton	S	29661	5.6	5.5
phosphatase 1, catalytic subunit, beta isoform 1	jgi	2538	Thalassiosira pseudonana	hydrolase	S	35512	5.0	4.7
Phosphoadenosine-phosphosulphate reductase	jgi	24887	Thalassiosira pseudonana	Metabolic Process	C	49035	5.0	3.9
Phosphofructokinase	jgi	22213	Thalassiosira pseudonana	Glycolysis	C	43797	5.7	2.7
Phosphoglucomutase, cytoplasmic (Glucose phosphomutase)	jgi	268621	Thalassiosira pseudonana	Binding Mg	S	60470	4.8	2.5
Phosphoglycerate kinase precursor	jgi	35712	Thalassiosira pseudonana	Glycolysis	C	42256	5.0	24.5
Phosphoglycerate mutase 1	jgi	27850	Thalassiosira pseudonana	Glycolysis	C	32465	6.1	4.1
Phosphoribulokinase	jgi	4376	Thalassiosira pseudonana	Biosynthesis	C	42389	4.9	7.2
Phosphoserine transaminase	jgi	3018	Thalassiosira pseudonana	Metabolic Process	C	55763	5.3	2.5
Photosystem I ferredoxin-binding protein	gi	118411153	Thalassiosira pseudonana	Photosynthesis	C	15518	9.6	56.8
Photosystem I protein F	gi	118411168	Thalassiosira pseudonana	Photosynthesis	C	20362	8.9	20.0
Photosystem I protein L	gi	118411163	Thalassiosira pseudonana	Photosynthesis	C	15704	9.3	5.4
Photosystem II 10 kDa phosphoprotein	gi	118411116	Thalassiosira pseudonana	Photosynthesis	C	7388	6.0	21.2
Photosystem II 11 kD protein	jgi	3258	Thalassiosira pseudonana	Photosystem	C	19602	9.6	6.3
Photosystem II chlorophyll A core	gi	118411113	Thalassiosira pseudonana	Photosynthesis	C	56408	6.5	14.5

antenna apoprotein								
Photosystem II chlorophyll A core antenna apoprotein CP43	gi	118411149	Thalassiosira pseudonana	Photosynthesis	C	51845	7.7	4.7
Photosystem II protein Y	gi	118411171	Thalassiosira pseudonana	Photosynthesis	C	4006	12.5	22.2
Photosystem II reaction center protein D1	gi	118411180	Thalassiosira pseudonana	Photosystem	C	39699	5.3	6.7
Photosystem II reaction center protein D2	gi	118411148	Thalassiosira pseudonana	Photosystem	C	39064	5.6	12.5
Photosystem II stability/assembly factor HCF136	jgi	38769	Thalassiosira pseudonana	Photosystem	C	40327	5.2	32.5
Phytanoyl-CoA dioxygenase	jgi	2770	Thalassiosira pseudonana	Oxygenase	S	33971	6.6	4.4
Phytoene dehydrogenase and related proteins	jgi	10233	Thalassiosira pseudonana	Electron Transport	C	75519	5.6	2.7
Polyprenyl synthetase	jgi	268480	Thalassiosira pseudonana	Biosynthesis	C	36015	4.8	4.1
Proteasomal ATPase	jgi	32037	Thalassiosira pseudonana	Binding DNA	C	45065	8.7	4.7
Protein product unnamed	jgi	37976	Thalassiosira pseudonana	Folding, Protein	U	17068	8.8	6.7
Pyridine nucleotide-disulphide oxidoreductase, class I	jgi	24399	Thalassiosira pseudonana	Electron Transport	S	52509	5.4	4.2
Pyruvate dehydrogenase E1 component beta subunit	jgi	32983	Thalassiosira pseudonana	Glycolysis	C	37156	5.4	3.5
Pyruvate kinase	jgi	4875	Thalassiosira pseudonana	Glycolysis	C	67221	5.2	4.5
Quinone oxidoreductase	jgi	32955	Thalassiosira pseudonana	Oxidation Reduction	C	32992	5.9	10.9
Ribosomal protein L12e	jgi	39424	Thalassiosira pseudonana	Translation	S	17412	9.1	14.6
Ribosomal protein L14	jgi	39499	Thalassiosira pseudonana	translation	C	14975	10.2	6.7
Ribosomal protein L19	jgi	268372	Thalassiosira pseudonana	Translation	S	21275	11.6	9.2
Ribosomal protein L5	jgi	802	Thalassiosira pseudonana	Translation	C	35283	8.6	3.9
Ribosomal protein PETRP-like	jgi	33241	Thalassiosira pseudonana	Translation	S	17754	10.3	8.1
Ribosomal protein S10	jgi	19501	Thalassiosira pseudonana	Translation	C	11912	9.4	8.7
Ribosomal protein S12	jgi	37628	Thalassiosira pseudonana	Binding RNA	C	12611	6.2	21.7
Ribosomal protein S18	jgi	26893	Thalassiosira pseudonana	Binding RNA	C	17159	10.8	28.8
Ribosomal protein S19	jgi	28425	Thalassiosira pseudonana	Translation	C	16703	9.0	8.8
Ribosomal protein S26e	jgi	20008	Thalassiosira pseudonana	Translation	S	10971	11.0	12.8
Ribosomal protein S3	jgi	28049	Thalassiosira pseudonana	Binding RNA	C	29161	9.3	9.6
Ribosomal protein S9	jgi	40312	Thalassiosira pseudonana	Translation	C	16088	10.2	5.6
Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	gi	118411104	Thalassiosira pseudonana	Photosynthesis	C	54325	6.2	35.1
Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	gi	118411103	Thalassiosira pseudonana	Photosynthesis	C	15843	5.1	69.8
Rieske iron-sulfur protein precursor	jgi	38231	Thalassiosira pseudonana	Electron Transport	C	19151	5.5	5.0
Rieske iron-sulfur protein precursor	jgi	26131	Thalassiosira pseudonana	Electron Transport	C	19305	5.1	8.8
RsuA	jgi	269764	Thalassiosira pseudonana	Binding RNA	S	36568	6.2	4.0

Rubisco expression protein	gi	118411164	Thalassiosira pseudonana	Photosynthesis	C	32381	5.9	22.3
S-adenosyl methionine synthetase	jgi	21815	Thalassiosira pseudonana	Transport	S	50359	5.2	16.1
S-adenosyl-L-homocysteinas protein	jgi	28496	Thalassiosira pseudonana	Metabolic Process	S	52309	5.1	10.6
S-adenosyl-L-homocysteine hydrolase	gi	71082903	Candidatus Pelagibacter ubique	Metabolism, one carbon	S	47071	5.5	4.0
Structural constituent of ribosome	jgi	26137	Thalassiosira pseudonana	Translation	S	20867	9.8	5.4
Structural constituent of ribosome	jgi	262056	Thalassiosira pseudonana	Translation	S	14755	10.4	9.7
Structural constituent of ribosome	jgi	31084	Thalassiosira pseudonana	Translation	S	13713	9.9	17.4
Transaldolase	jgi	27187	Thalassiosira pseudonana	Metabolic Process Carbohydrate	S	34855	4.8	11.0
Translation elongation factor G	jgi	25629	Thalassiosira pseudonana	Translation	C	86389	5.0	16.6
Translation elongation factor Tu	gi	118411218	Thalassiosira pseudonana	Translation	C	44458	4.9	28.9
Triosephosphate isomerase/glyceraldehyde-3-phosphate dehydrogenase precursor	jgi	28239	Thalassiosira pseudonana	Metabolic Process	S	65308	5.6	4.9
Tubulin alpha-2 chain	jgi	29304	Thalassiosira pseudonana	Structural	S	49904	5.0	8.4
Tubulin beta chain	jgi	8069	Thalassiosira pseudonana	Structural	S	49497	4.9	3.3
Tubulin beta chain	jgi	31569	Thalassiosira pseudonana	Structural	S	49670	4.9	12.6
Ubiquinol-cytochrome-c reductase	jgi	36107	Thalassiosira pseudonana	Oxidoreductase	M	11977	5.4	9.3
Ubiquitin	jgi	40669	Thalassiosira pseudonana	Modification Protein	S	17567	9.9	18.3
Unknown	jgi	10417	Thalassiosira pseudonana	n.a.	S	22939	4.9	6.4
Unknown	jgi	30683	Thalassiosira pseudonana	Metabolic Process	U	26039	6.0	6.3
Unknown	jgi	39424	Thalassiosira pseudonana	Binding GTP	U	20905	6.8	25.6
Vacuolar ATP synthase 16 kDa proteolipid subunit	jgi	2233	Thalassiosira pseudonana	Transport Proton	S	16720	5.6	10.8
Vacuolar ATP synthase subunit A	jgi	37123	Thalassiosira pseudonana	Transport Proton	S	68343	5.0	7.6
Vacuolar ATPase B subunit	jgi	40522	Thalassiosira pseudonana	Transport Proton	S	56064	5.9	16.8
Vacuolar proton-inorganic pyrophosphatase	jgi	39520	Thalassiosira pseudonana	Transport Proton	S	70120	5.0	2.7
Vacuolar sorting receptor homolog	jgi	42545	Thalassiosira pseudonana	Binding Calcium	S	56362	4.9	3.1

50m POC

Protein	Annotation		Species	Function	Comp	MW	pI	SC
ABC transporter	gi	71083646	Candidatus Pelagibacter ubique	Transport	S	25649	6.1	4
Actin A	jgi	25772	Thalassiosira pseudonana	Binding Protein	S	41791	5.0	2.7
Adenosine-triphosphatase	jgi	40156	Thalassiosira pseudonana	Transport Proton	S	39935	7.6	3.5
ATP synthase CF1 alpha chain	gi	118411112	Thalassiosira pseudonana	Transport Ion	C	53989	5.0	4.6
ATP synthase CF1 beta	gi	118411134	Thalassiosira	Transport Ion	C	51143	4.7	6.1

chain			pseudonana					
Cytochrome c-550	gi	118411100	Thalassiosira pseudonana	Electron Transport	C	17841	7.7	11
Fucoxanthin chlorophyll a/c binding protein	jgi	268127	Thalassiosira pseudonana	Photosynthesis	C	22628	4.8	14.8
Glyceraldehyde-3-phosphate dehydrogenase precursor	jgi	31383	Thalassiosira pseudonana	Glycolysis	C	39587	5.3	4
Hypothetical Protein	jgi	23918	Thalassiosira pseudonana	n.a.	S	31192	4.8	3.1
Ribulose-1,5-bisphosphate carboxylase/ oxygenase large subunit	gi	118411104	Thalassiosira pseudonana	Photosynthesis	C	54325	6.2	3.9
Tubulin beta chain	jgi	8069	Thalassiosira pseudonana	Structural	S	49497	4.9	2.7

100m POC

Protein	Annotation		Species	Function	Comp	MW	pI	SC
ABC transporter	gi	71083646	Candidatus Pelagibacter ubique	Transport	S	25649	6.1	6.6
Adenosinetriphosphatase	jgi	40156	Thalassiosira pseudonana	Transport Proton	S	39935	7.6	3.5
ATP synthase CF1 alpha chain	gi	118411112	Thalassiosira pseudonana	Transport Ion	C	53989	5.0	5
ATP synthase CF1 beta chain	gi	118411134	Thalassiosira pseudonana	Transport Ion	C	51143	4.7	8.4
CG11154-PA, isoform A	jgi	41256	Thalassiosira pseudonana	Transport Proton	U	53388	5.1	3
DNA binding	jgi	29950	Thalassiosira pseudonana	Binding DNA	C	15300	11.3	13.2
Fucoxanthin chlorophyl a/c protein	jgi	38667	Thalassiosira pseudonana	Photosynthesis	C	21807	4.8	6.8
Fucoxanthin chlorophyll a/c binding protein	jgi	268127	Thalassiosira pseudonana	Photosynthesis	C	22628	4.8	6.2
Glyceraldehyde-3-phosphate dehydrogenase precursor	jgi	31383	Thalassiosira pseudonana	Glycolysis	S	39587	5.3	4
Histone H2A.1	jgi	19793	Thalassiosira pseudonana	Binding DNA	N	13053	10.4	7.3
Histone H4	jgi	3184	Thalassiosira pseudonana	Binding DNA	N	11384	11.5	34
Hypothetical protein DEHA0F19712g	jgi	27352	Thalassiosira pseudonana	Binding DNA	N	13064	10.3	12.5
Photosystem II chlorophyll A core antenna apoprotein	gi	118411113	Thalassiosira pseudonana	Photosynthesis	C	56408	6.5	4.1
Photosystem II reaction center protein D2	gi	118411148	Thalassiosira pseudonana	Photosynthesis	C	39064	5.6	6.3
Protein product unnamed	jgi	27435	Thalassiosira pseudonana	Transcription	U	18469	6.3	7.7
Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	gi	118411104	Thalassiosira pseudonana	Photosynthesis	C	54325	6.2	16.3
Spermidine/putrescine-binding periplasmic protein	gi	71084024	Candidatus Pelagibacter ubique	Binding	S	40556	5.0	8.8
TRAP dicarboxylate transporter - DctP subunit	gi	71082971	Candidatus Pelagibacter ubique	Transport	S	42131	9.2	6.2
Tubulin alpha-2 chain	jgi	29304	Thalassiosira pseudonana	Structural	S	49904	5.0	5.3
Tubulin beta chain	jgi	8069	Thalassiosira pseudonana	Structural	S	49497	4.9	12.9

Ubiquitin	jgi	40669	Thalassiosira pseudonana	Modification Protein	S	17567	9.9	8.5
Vacuolar-type H+-pyrophosphatase	jgi	32586	Thalassiosira pseudonana	Transport Proton	S	67969	4.6	1.7

40m Trap

Protein	Annotation		Species	Function	Comp	MW	pI	SC
30S ribosomal protein S1	jgi	15259	Thalassiosira pseudonana	Binding RNA	C	31770	4.6	7.3
30S ribosomal protein S12	gi	118411216	Thalassiosira pseudonana	Translation	C	13914	11.6	6.5
30S ribosomal protein S18	gi	118411132	Thalassiosira pseudonana	Translation	C	8155	10.8	18.1
3-deoxy-7-phosphoheptulonate synthase	jgi	2790	Thalassiosira pseudonana	Biosynthesis	S	53939	6.0	9.9
40S ribosomal protein S5	jgi	29955	Thalassiosira pseudonana	Translation	R	24397	6.9	6
40S ribosomal protein S9	jgi	268651	Thalassiosira pseudonana	Binding rRNA	R	21764	10.2	11.6
40S ribosomal protein SA p40	jgi	21871	Thalassiosira pseudonana	Translation	R	27261	5.9	13.2
50S ribosomal protein L4	gi	118411191	Thalassiosira pseudonana	Translation	C	24203	10.2	7.9
50S ribosomal protein L5	gi	118411203	Thalassiosira pseudonana	Translation	C	27571	9.7	5.9
60 kDa chaperonin	gi	118411188	Thalassiosira pseudonana	Binding Protein	C	57361	5.2	5.3
6-phosphogluconate dehydrogenase	jgi	33343	Thalassiosira pseudonana	Dehydrogenase	S	53348	5.6	8.4
Acidic ribosomal phosphoprotein P0	jgi	25812	Thalassiosira pseudonana	Biosynthesis Ribosome	R	34116	4.8	3.7
Actin A	jgi	25772	Thalassiosira pseudonana	Binding Protein	S	41791	5.0	15.1
Adenosinetriphosphatase	jgi	40156	Thalassiosira pseudonana	Transport Proton	S	39935	7.6	7.4
Aminotransferase AGD2	jgi	31394	Thalassiosira pseudonana	Transport	M	44264	4.8	3.7
Argininosuccinate synthase	jgi	42719	Thalassiosira pseudonana	Biosynthesis	C	45938	5.3	2.9
Aromatic-ring hydroxylase	gi	33861317	Prochlorococcus marinus	Metabolic Process	S	49438	8.9	3.1
ATP binding / protein binding	jgi	23102	Thalassiosira pseudonana	Binding Protein	S	59170	5.3	5.3
ATP synthase CF0 B chain subunit I	gi	118411110	Thalassiosira pseudonana	Transport Proton	C	20029	9.8	16.8
ATP synthase CF0 B' chain subunit II	gi	118411109	Thalassiosira pseudonana	Transport Proton	C	17373	4.6	9.6
ATP synthase CF1 alpha chain	gi	118411112	Thalassiosira pseudonana	Transport Ion	C	53989	5.0	25.8
ATP synthase CF1 beta chain	gi	118411134	Thalassiosira pseudonana	Transport Ion	C	51143	4.7	36.5
ATP/ADP translocator	jgi	39143	Thalassiosira pseudonana	Transport	M	32254	9.4	3.6
ATP-dependent clp protease ATP-binding subunit	gi	118411220	Thalassiosira pseudonana	Catalysis	C	102150	6.5	1.5
ATP-dependent clp protease ATP-binding subunit	gi	33861644	Prochlorococcus marinus	n.a.	S	93370	5.5	2.6
ATP-sulfurylase	jgi	1326	Thalassiosira pseudonana	Metabolic Process	M	45362	5.2	4.9
BiP	jgi	27656	Thalassiosira pseudonana	Cell Morphogenesis	S	70451	4.7	20
CbbX protein homolog	jgi	40193	Thalassiosira	Biosynthesis	C	35036	5.3	19.3

			pseudonana					
CDC48/ATPase	jgi	267952	Thalassiosira pseudonana	Binding ATP	S	89464	4.8	3.2
Cell division protein FtsH2	jgi	31930	Thalassiosira pseudonana	Binding Zn	C	61956	5.1	16
Cell division protein FtsH-like protein	gi	118411141	Thalassiosira pseudonana	Proteolysis	C	70206	5.1	9.8
CG11154-PA, isoform A	jgi	41256	Thalassiosira pseudonana	Transport Proton	U	53388	5.1	29
Chloroplast 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase precursor	jgi	29228	Thalassiosira pseudonana	Biosynthesis	C	75736	4.9	2.6
Chloroplast ferredoxin dependent NADH oxireductase	jgi	25892	Thalassiosira pseudonana	Transport Electron	C	37819	5.9	2.4
Chloroplast light harvesting protein isoform 12	jgi	270092	Thalassiosira pseudonana	Photosynthesis	C	26078	5.5	11.2
Chloroplast light harvesting protein isoform 15	jgi	2845	Thalassiosira pseudonana	Photosynthesis	C	21873	5.1	5.4
Chorismate synthase	jgi	38964	Thalassiosira pseudonana	Biosynthesis	S	44051	5.5	3.2
Cobaltochelatase	jgi	26573	Thalassiosira pseudonana	Biosynthesis	S	148275	4.9	1
CPN60 protein	jgi	23329	Thalassiosira pseudonana	Binding Protein	M	59177	4.7	5.1
Cytochrome b559 alpha chain	gi	118411160	Thalassiosira pseudonana	Photosynthesis	C	9514	5.6	25
Cytochrome c-550	gi	118411100	Thalassiosira pseudonana	Transport Electron	C	17841	7.7	11
Cytochrome f	gi	118411137	Thalassiosira pseudonana	Photosynthesis	C	33988	8.2	19.1
DNA binding	jgi	29950	Thalassiosira pseudonana	Binding DNA	C	15300	11.3	8.1
Domain specific binding protein 14-3-3	jgi	26146	Thalassiosira pseudonana	Binding Protein Domain Specific	S	27869	4.6	7.3
Elongation factor 2	jgi	269148	Thalassiosira pseudonana	Translation	S	91887	6.0	4.2
Enolase 2	jgi	40391	Thalassiosira pseudonana	Glycolysis	S	46547	4.8	3.4
Enoyl-acyl carrier reductase	jgi	32860	Thalassiosira pseudonana	Oxidation Reduction	S	32813	5.1	8
Eukaryotic translation initiation factor 4A2 isoform 2	jgi	9716	Thalassiosira pseudonana	Binding Nucleic Acid	S	42405	5.6	20.6
F0F1 ATP synthase subunit alpha	gi	33862007	Prochlorococcus marinus	Transport Proton	S	54306	4.9	6.7
F0F1 ATP synthase subunit beta	gi	71082935	Candidatus Pelagibacter ubique	Transport Proton	S	50696	4.9	16.9
FeS assembly protein SufD	jgi	268364	Thalassiosira pseudonana	Binding Protein	S	31963	5.2	5.8
Fructose-1,6-bisphosphate aldolase precursor	jgi	428	Thalassiosira pseudonana	Glycolysis	S	39810	4.8	7.3
Fructose-bisphosphatase	jgi	264556	Thalassiosira pseudonana	Metabolic Process Carbohydrate	M	33667	5.3	5.1
Fucoxanthin chlorophyl a/c protein	jgi	38667	Thalassiosira pseudonana	Photosynthesis	C	21807	4.8	21.6
Fucoxanthin chlorophyll a /c protein	jgi	38494	Thalassiosira pseudonana	Photosynthesis	C	20354	4.5	18.4
Fucoxanthin chlorophyll a /c protein	jgi	42962	Thalassiosira pseudonana	Photosynthesis	C	21515	5.1	17.5
Fucoxanthin chlorophyll a/c binding protein	jgi	264921	Thalassiosira pseudonana	Photosynthesis	C	21263	5.0	5
Fucoxanthin chlorophyll	jgi	268127	Thalassiosira	Photosynthesis	C	22205	4.6	15.3

a/c binding protein			pseudonana					
Fucoxanthin chlorophyll a/c binding protein	jgi	30385	Thalassiosira pseudonana	Photosynthesis	C	22628	4.8	35.7
GDP-mannose dehydratase	jgi	40586	Thalassiosira pseudonana	Catalysis	S	40412	5.9	3.3
Geranyl-geranyl reductase	jgi	10234	Thalassiosira pseudonana	Transport Electron	C	47230	5.9	7.3
Glucose-6-phosphate isomerase	jgi	38266	Thalassiosira pseudonana	Glycolysis	C	61689	5.9	4.9
Glutamate 1-semialdehyde 2,1-aminomutase	jgi	575	Thalassiosira pseudonana	Binding Phosphate	C	43658	5.5	3.2
Glutamine synthase	jgi	270138	Thalassiosira pseudonana	Biosynthesis	C	69172	5.2	1.8
Glutamine synthetase	jgi	26051	Thalassiosira pseudonana	Biosynthesis	C	45620	5.2	3.4
Glyceraldehyde-3-phosphate dehydrogenase	jgi	28334	Thalassiosira pseudonana	Glycolysis	M	36574	6.1	4.4
Glyceraldehyde-3-phosphate dehydrogenase precursor	jgi	31383	Thalassiosira pseudonana	Glycolysis	S	39587	5.3	27.2
Glycolaldehydetransferase	jgi	21175	Thalassiosira pseudonana	Transport	M	71708	5.0	26.1
Heat shock protein 60	jgi	38191	Thalassiosira pseudonana	Binding Protein	S	58525	4.8	2.9
Heat shock protein 70	jgi	269120	Thalassiosira pseudonana	Folding Protein	S	71187	4.8	16.4
Heat shock protein Hsp90	jgi	6285	Thalassiosira pseudonana	Folding Protein	S	80242	4.7	7.1
Histone H2A.1	jgi	19793	Thalassiosira pseudonana	Binding DNA	N	13053	10.4	7.3
Histone H4	jgi	3184	Thalassiosira pseudonana	Binding DNA	N	11384	11.5	35
Hsp70-type chaperone	gi	118411189	Thalassiosira pseudonana	Transcription	C	65339	4.8	2.3
Hypothetical Protein	jgi	23918	Thalassiosira pseudonana	n.a.	S	31192	4.8	36.5
Hypothetical protein CBG01077	jgi	22792	Thalassiosira pseudonana	Transport	S	22486	6.7	9
Hypothetical protein CBG08717	jgi	269322	Thalassiosira pseudonana	Transport Proton	S	58009	5.8	19.4
Hypothetical protein FG01081.1	jgi	25949	Thalassiosira pseudonana	Translation	U	20124	9.8	7.9
Hypothetical protein Synpcc7942_1497	jgi	4382	Thalassiosira pseudonana	n.a.	S	30101	5.6	5.3
Hypothetical protein UM03322.1	jgi	28443	Thalassiosira pseudonana	Translation	M	24676	10.2	5.1
Isocitrate/isopropylmalate dehydrogenase	jgi	5293	Thalassiosira pseudonana	Oxidoreductase	S	40667	4.6	5.9
Ketol-acid reductoisomerase	jgi	23228	Thalassiosira pseudonana	Oxidation Reduction	S	58240	5.1	7.3
L4/L1	jgi	22610	Thalassiosira pseudonana	Translation	R	40991	10.3	5
Malate dehydrogenase	jgi	20726	Thalassiosira pseudonana	Oxidoreductase	M	36724	6.3	3.4
Mitochondrial glyceraldehyde-3-phosphate dehydrogenase	jgi	28241	Thalassiosira pseudonana	Glycolysis	M	36243	5.9	9.5
Molecular chaperone DnaK2	gi	33862260	Prochlorococcus marinus	Folding Protein	S	68202	4.8	4.9
Myo-inositol dehydrogenase precursor	jgi	1049	Thalassiosira pseudonana	Biosynthesis	S	48134	5.0	4.2
Nitrate reductase	jgi	25299	Thalassiosira pseudonana	Transport Electron	S	101408	5.9	1
Oxidoreductase	jgi	8063	Thalassiosira pseudonana	Metabolic Process	C	36723	7.6	6.2
Oxygen-evolving enhancer protein 1 precursor	jgi	34830	Thalassiosira pseudonana	Photosynthesis	C	29136	5.2	26.2

Phosphoadenosine-phosphosulphate reductase	jgi	24887	Thalassiosira pseudonana	Metabolic Process	C	49035	5.0	3.9
Phosphofructokinase	jgi	22213	Thalassiosira pseudonana	Glycolysis	C	43797	5.7	2.7
Phosphoglycerate kinase precursor	jgi	35712	Thalassiosira pseudonana	Glycolysis	S	42256	5.0	26.8
Phosphoglycerate mutase 1	jgi	27850	Thalassiosira pseudonana	Glycolysis	C	32465	6.1	3.1
Phosphoribosyl-pyrophosphate synthetase	jgi	26109	Thalassiosira pseudonana	Biosynthesis	S	33425	8.4	3.2
Photosystem I ferredoxin-binding protein	gi	118411153	Thalassiosira pseudonana	Photosynthesis	C	15518	9.6	37.4
Photosystem I p700 chlorophyll A apoprotein A	gi	118411096	Thalassiosira pseudonana	Photosynthesis	C	83642	7.3	4.1
Photosystem I protein F	gi	118411168	Thalassiosira pseudonana	Photosynthesis	C	20362	8.9	8.1
Photosystem II 10 kDa phosphoprotein	gi	118411116	Thalassiosira pseudonana	Photosynthesis	C	7388	6.0	21.2
Photosystem II chlorophyll A core antenna apoprotein	gi	118411113	Thalassiosira pseudonana	Photosynthesis	C	56408	6.5	20.2
Photosystem II chlorophyll A core antenna apoprotein CP43	gi	118411149	Thalassiosira pseudonana	Photosynthesis	C	51845	7.7	12.1
Photosystem II PsbD protein D2	gi	33861713	Prochlorococcus marinus	Photosynthesis	C	39917	5.6	6.4
Photosystem II reaction center protein D1	gi	118411180	Thalassiosira pseudonana	Photosystem	C	39699	5.3	6.4
Photosystem II reaction center protein D2	gi	118411148	Thalassiosira pseudonana	Photosystem	C	39064	5.6	12.5
Photosystem II stability/assembly factor HCF136	jgi	38769	Thalassiosira pseudonana	Photosystem	C	40327	5.2	10.5
Protein product unnamed	jgi	262083	Thalassiosira pseudonana	n.a.	S	103903	6.1	2.6
Pyruvate kinase	jgi	40393	Thalassiosira pseudonana	Glycolysis	C	54650	5.9	2.5
Ribosomal protein L12e	jgi	39424	Thalassiosira pseudonana	Translation	R	17412	9.1	14.6
Ribosomal protein S13	jgi	26221	Thalassiosira pseudonana	Translation	C	17054	10.4	8.6
Ribosomal protein S18	jgi	26893	Thalassiosira pseudonana	Binding RNA	C	17159	10.8	27.4
Ribosomal protein S3	jgi	28049	Thalassiosira pseudonana	Binding RNA	C	29161	9.3	8.9
Ribosomal protein S9	jgi	40312	Thalassiosira pseudonana	Translation	C	16088	10.2	7.7
Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	gi	118411104	Thalassiosira pseudonana	Photosynthesis	C	54325	6.2	32.2
Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	gi	118411103	Thalassiosira pseudonana	Photosynthesis	C	15843	5.1	19.4
RsuA	jgi	269764	Thalassiosira pseudonana	Binding RNA	S	36568	6.2	4
Rubisco expression protein	gi	118411164	Thalassiosira pseudonana	Photosynthesis	C	32381	5.9	4.9
S-adenosyl methionine synthetase	jgi	21815	Thalassiosira pseudonana	Transport	S	50359	5.2	6
S-adenosyl-L-homocysteinas protein	jgi	28496	Thalassiosira pseudonana	Metabolic Process	S	52309	5.1	12.7
S-adenosyl-L-homocysteine hydrolase	gi	71082903	Candidatus Pelagibacter ubique	Metabolic Process	S	47071	5.5	4
SDH1-1	jgi	42475	Thalassiosira	Transport	M	68963	5.7	2.2

			pseudonana	Electron				
Serine hydroxymethyltransferase	jgi	26031	Thalassiosira pseudonana	Metabolic Process	S	52999	6.5	1.9
Transaldolase	jgi	27187	Thalassiosira pseudonana	Metabolic Process Carbohydrate	S	34855	4.8	12.6
Translation elongation factor 1 alpha	jgi	3858	Thalassiosira pseudonana	Translation	S	47815	7.2	2.5
Translation elongation factor G	jgi	25629	Thalassiosira pseudonana	Translation	S	86389	5.0	9.2
Translation elongation factor Tu	gi	118411218	Thalassiosira pseudonana	Translation	C	44458	4.9	17.8
Tubulin alpha-2 chain	jgi	29304	Thalassiosira pseudonana	Structure	S	49904	5.0	11.9
Tubulin beta chain	jgi	31569	Thalassiosira pseudonana	Structure	S	49497	4.9	18.5
Tubulin beta chain	jgi	8069	Thalassiosira pseudonana	Structure	S	49670	4.9	3.3
Ubiquitin	jgi	40669	Thalassiosira pseudonana	Modification Protein	S	17567	9.9	8.5
Unknown	jgi	10417	Thalassiosira pseudonana	n.a.	S	22939	4.9	17.8
Vacuolar ATP synthase 16 kDa proteolipid subunit	jgi	2233	Thalassiosira pseudonana	Transport Proton	S	16720	5.6	10.8
Vacuolar ATP synthase subunit A	jgi	37123	Thalassiosira pseudonana	Transport Proton	S	68343	5.0	3.5
Vacuolar ATPase B subunit	jgi	40522	Thalassiosira pseudonana	Transport Proton	S	56064	5.9	23.9
Vacuolar proton translocating ATPase A subunit	jgi	40728	Thalassiosira pseudonana	Transport Proton	S	92414	5.3	1.6
Vacuolar-type H+-pyrophosphatase	jgi	32586	Thalassiosira pseudonana	Transport Proton	S	67969	4.6	1.7
Vitamin B6 biosynthesis protein	jgi	42612	Thalassiosira pseudonana	Biosynthesis	S	35430	6.0	9

60m Trap

Protein	Annotation		Species	Function	Comp	MW	pI	SC
30S ribosomal protein S1	jgi	15259	Thalassiosira pseudonana	Translation	R	31770	4.6	3.8
3-deoxy-7-phosphoheptulonate synthase	jgi	2790	Thalassiosira pseudonana	Biosynthesis	C	53939	6.0	4.8
3-phosphoshikimate 1-carboxyvinyltransferase	jgi	33008	Thalassiosira pseudonana	Transport	C	47333	4.6	3.1
6-phosphogluconate dehydrogenase	jgi	33343	Thalassiosira pseudonana	Dehydrogenase	S	53348	5.6	2
Actin A	jgi	25772	Thalassiosira pseudonana	Binding Protein	S	41791	5.0	9
Adenosinetriphosphatase	jgi	40156	Thalassiosira pseudonana	Transport Proton	S	39935	7.6	3.5
Aromatic-ring hydroxylase	gi	33861317	Prochlorococcus marinus	Photosynthesis	S	49438	8.9	3.1
ATP binding / protein binding	jgi	23102	Thalassiosira pseudonana	Binding Protein	S	59170	5.3	5.3
ATP synthase CF1 alpha chain	gi	118411112	Thalassiosira pseudonana	Transport Ion	C	53989	5.0	24.1
ATP synthase CF1 beta chain	gi	118411134	Thalassiosira pseudonana	Transport Ion	C	51143	4.7	38.4
ATP-sulfurylase	jgi	1326	Thalassiosira pseudonana	Metabolic Process	M	45362	5.2	2.9
BiP	jgi	27656	Thalassiosira pseudonana	Cell Morphogenesis	S	70451	4.7	4.5
CbbX protein homolog	jgi	40193	Thalassiosira	Biosynthesis	C	35036	5.3	7.1

			pseudonana					
CG11154-PA, isoform A	jgi	41256	Thalassiosira pseudonana	Transport Proton	S	53388	5.1	33
CPN60 protein	jgi	23329	Thalassiosira pseudonana	Binding Protein	M	59177	4.7	2.9
Enolase 2	jgi	40391	Thalassiosira pseudonana	Glycolysis	S	46547	4.8	5.5
Eukaryotic translation initiation factor 4A2 isoform 2	jgi	9716	Thalassiosira pseudonana	Binding DNA	S	42405	5.6	17.6
Fructose-1,6-bisphosphate aldolase precursor	jgi	428	Thalassiosira pseudonana	Glycolysis	S	39810	4.8	12.4
Fructose-bisphosphatase	jgi	264556	Thalassiosira pseudonana	Metabolic Process Carbohydrate	M	33667	5.3	5.1
Fucoxanthin chlorophyl a/c protein	jgi	38667	Thalassiosira pseudonana	Photosynthesis	C	21807	4.8	13.2
Fucoxanthin chlorophyll a /c protein	jgi	38494	Thalassiosira pseudonana	Photosynthesis	C	20354	4.5	13.7
Fucoxanthin chlorophyll a/c binding protein	jgi	264921	Thalassiosira pseudonana	Photosynthesis	C	22205	4.6	8.6
Fucoxanthin chlorophyll a/c binding protein	jgi	268127	Thalassiosira pseudonana	Photosynthesis	C	22628	4.8	27.6
Glutamine synthetase	jgi	26051	Thalassiosira pseudonana	Biosynthesis	C	45620	5.2	3.4
Glyceraldehyde-3-phosphate dehydrogenase	jgi	28334	Thalassiosira pseudonana	Glycolysis	S	36574	6.1	4.4
Glyceraldehyde-3-phosphate dehydrogenase precursor	jgi	31383	Thalassiosira pseudonana	Glycolysis	S	39587	5.3	10.4
Glycolaldehydetransferase	jgi	21175	Thalassiosira pseudonana	Transport	S	71708	5.0	4.4
Heat shock protein 60	jgi	38191	Thalassiosira pseudonana	Folding Protein	S	58525	4.8	5.2
Heat shock protein 70	jgi	269120	Thalassiosira pseudonana	Folding Protein	S	71187	4.8	10.3
Histone H2A.1	jgi	19793	Thalassiosira pseudonana	Binding DNA	N	13053	10.4	7.3
Hypothetical Protein	jgi	23918	Thalassiosira pseudonana	n.a.	S	31192	4.8	9.6
Hypothetical protein CBG08717	jgi	269322	Thalassiosira pseudonana	Transport Proton	S	58009	5.8	12.9
Isocitrate/isopropylmalate dehydrogenase	jgi	5293	Thalassiosira pseudonana	Oxidoreductase	S	40667	4.6	6.4
L4/L1	jgi	22610	Thalassiosira pseudonana	Translation	R	40991	10.3	4.2
Magnesium-chelatase subunit I	gi	118411138	Thalassiosira pseudonana	Photosynthesis	C	39500	5.0	5.1
Mitochondrial glyceraldehyde-3-phosphate dehydrogenase	jgi	28241	Thalassiosira pseudonana	Glycolysis	M	36243	5.9	6.2
Phosphoadenosine-phosphosulphate reductase	jgi	24887	Thalassiosira pseudonana	Metabolic Process	C	49035	5.0	3.9
Phosphoglycerate kinase precursor	jgi	35712	Thalassiosira pseudonana	Glycolysis	S	42256	5.0	15
Photosystem II 10 kDa phosphoprotein	gi	118411116	Thalassiosira pseudonana	Photosynthesis	C	7388	6.0	21.2
Photosystem II chlorophyll A core antenna apoprotein	gi	118411113	Thalassiosira pseudonana	Photosynthesis	C	56408	6.5	5.5
Photosystem II chlorophyll A core antenna apoprotein CP43	gi	118411149	Thalassiosira pseudonana	Photosynthesis	C	51845	7.7	3.4
Photosystem II reaction center protein D2	gi	118411148	Thalassiosira pseudonana	Photosystem	C	39064	5.6	6.3
Photosystem II stability/assembly factor	jgi	38769	Thalassiosira pseudonana	Photosystem	C	40327	5.2	7.3

HCF136								
Predicted translation elongation factor G	jgi	25629	Thalassiosira pseudonana	Translation	S	86389	5.0	6.6
Putative aminotransferase AGD2	jgi	31394	Thalassiosira pseudonana	Transport	M	44264	4.8	6.6
Putative S-adenosyl-L-homocysteinas protein	jgi	28496	Thalassiosira pseudonana	Metabolic Process	S	52309	5.1	6.2
Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	gi	118411104	Thalassiosira pseudonana	Photosynthesis	C	54325	6.2	12.2
Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	gi	118411103	Thalassiosira pseudonana	Photosynthesis	C	15843	5.1	10.1
S-adenosyl methionine synthetase	jgi	21815	Thalassiosira pseudonana	Transport	S	50359	5.2	3.2
Transaldolase	jgi	27187	Thalassiosira pseudonana	Metabolic Process Carbohydrate	S	34855	4.8	4.1
Translation elongation factor Tu	gi	118411218	Thalassiosira pseudonana	Translation	C	44458	4.9	3.9
Tubulin alpha-2 chain	jgi	29304	Thalassiosira pseudonana	Structural	S	49904	5.0	2.2
Vacuolar ATPase B subunit	jgi	40522	Thalassiosira pseudonana	Transport Proton	S	56064	5.9	12.8

100m Trap

Protein	Annotation		Species	Function	Comp	MW	pI	SC
30S ribosomal protein S12	gi	118411216	Thalassiosira pseudonana	Translation	C	13914	11.6	6.5
3-deoxy-7-phosphoheptulonate synthase	jgi	2790	Thalassiosira pseudonana	Transport Electron	C	53939	6.0	1.9
40S ribosomal protein S17-like protein	jgi	37809	Thalassiosira pseudonana	Translation	C	14145	10.0	17.1
40S ribosomal protein S6	jgi	269779	Thalassiosira pseudonana	Translation	R	27587	11.0	3.7
40S ribosomal protein SA p40	jgi	21871	Thalassiosira pseudonana	Translation	R	27261	5.9	13.2
50S ribosomal protein L11	gi	118411123	Thalassiosira pseudonana	Translation	C	14880	9.7	9.2
50S ribosomal protein L4	gi	118411191	Thalassiosira pseudonana	Translation	C	24203	10.2	5.1
6-phosphogluconate dehydrogenase	jgi	33343	Thalassiosira pseudonana	Dehydrogenase	S	53348	5.6	6.9
Actin A	jgi	25772	Thalassiosira pseudonana	Binding Protein	S	41791	5.0	11.7
Adenosinetriphosphatase	jgi	40156	Thalassiosira pseudonana	Transport Proton	S	39935	7.6	3.5
Aromatic-ring hydroxylase	gi	33861317	Prochlorococcus marinus	Metabolic Process	S	49438	8.9	3.1
ATP synthase CF1 alpha chain	gi	118411112	Thalassiosira pseudonana	Transport Ion	C	53989	5.0	14.3
ATP synthase CF1 beta chain	gi	118411134	Thalassiosira pseudonana	Transport Ion	C	51143	4.7	14.1
ATPase, E1-E2 type	jgi	262679	Thalassiosira pseudonana	Transport Cation	S	99192	5.6	1.5
ATP-dependent clp protease ATP-binding subunit	gi	118411220	Thalassiosira pseudonana	Catalysis	C	102150	6.5	1.5
BiP	jgi	27656	Thalassiosira pseudonana	Cell Morphogenesis	S	70451	4.7	6.7
Cell division protein FtsH-like protein	gi	118411141	Thalassiosira pseudonana	Proteolysis	C	70206	5.1	1.9

CG11154-PA, isoform A	jgi	41256	Thalassiosira pseudonana	Transport Proton	S	53388	5.1	11.6
Chloroplast ferredoxin dependent NADH oxireductase	jgi	25892	Thalassiosira pseudonana	Transport Electron	C	37819	5.9	4.4
Cyc07-like protein	jgi	26046	Thalassiosira pseudonana	Translation	S	28811	9.6	13.1
Cytochrome b559 alpha chain	gi	118411160	Thalassiosira pseudonana	Photosynthesis	C	9514	5.6	10.7
Cytochrome b6	gi	118411154	Thalassiosira pseudonana	Transport Electron	C	23906	9.2	6
Cytochrome c-550	gi	118411100	Thalassiosira pseudonana	Transport Electron	C	17841	7.7	11
Cytochrome f	gi	118411137	Thalassiosira pseudonana	Photosynthesis	C	33988	8.2	4.1
DNA binding	jgi	29950	Thalassiosira pseudonana	Binding DNA	C	15300	11.3	13.2
Elongation factor alpha-like protein	jgi	41829	Thalassiosira pseudonana	Translation	S	49969	8.7	2.2
Eukaryotic translation initiation factor 4A2 isoform 2	jgi	9716	Thalassiosira pseudonana	Binding Nucleic Acid	S	42405	5.6	6
Fructose-1,6-bisphosphate aldolase precursor	jgi	428	Thalassiosira pseudonana	Glycolysis	S	39810	4.8	3
Fucoxanthin chlorophyll a /c protein	jgi	38494	Thalassiosira pseudonana	Photosynthesis	C	20354	4.5	11.6
Fucoxanthin chlorophyll a/c binding protein	jgi	264921	Thalassiosira pseudonana	Photosynthesis	C	22205	4.6	8.6
Fucoxanthin chlorophyll a/c binding protein	jgi	268127	Thalassiosira pseudonana	Photosynthesis	C	22628	4.8	11
Fucoxanthin-chlorophyll a/c light-harvesting protein	jgi	33018	Thalassiosira pseudonana	Photosynthesis	C	21786	5.4	4.5
GDP-mannose dehydratase	jgi	40586	Thalassiosira pseudonana	Catalysis	S	40412	5.9	3.3
Glutamine synthetase	jgi	26051	Thalassiosira pseudonana	Biosynthesis	C	45620	5.2	5.8
Glyceraldehyde-3-phosphate dehydrogenase	jgi	28334	Thalassiosira pseudonana	Glycolysis	M	36574	6.1	4.4
Glyceraldehyde-3-phosphate dehydrogenase precursor	jgi	31383	Thalassiosira pseudonana	Glycolysis	S	39587	5.3	8.8
Glycolaldehydetransferase	jgi	21175	Thalassiosira pseudonana	Transport	S	71708	5.0	4.4
Heat shock protein 70	jgi	269120	Thalassiosira pseudonana	Folding Protein	S	71187	4.8	2
Heat shock protein Hsp90	jgi	6285	Thalassiosira pseudonana	Folding Protein	S	80242	4.7	2
Histone H2A.1	jgi	19793	Thalassiosira pseudonana	Binding DNA	N	13053	10.4	7.3
Histone H4	jgi	3184	Thalassiosira pseudonana	Binding DNA	N	11384	11.5	30.1
Hypothetical protein	jgi	27167	Thalassiosira pseudonana	n.a.	U	27052	10.0	5.4
Hypothetical Protein	jgi	23918	Thalassiosira pseudonana	n.a.	S	31192	4.8	33.1
Hypothetical protein CBG08717	jgi	269322	Thalassiosira pseudonana	Transport Proton	S	58009	5.8	6.5
hypothetical protein FG01081.1	jgi	25949	Thalassiosira pseudonana	Translation	U	20124	9.8	7.9
Hypothetical protein UM03322.1	jgi	28443	Thalassiosira pseudonana	Translation	R	24676	10.2	10.6
Inorganic diphosphatase/ magnesium ion binding / pyrophosphatase	jgi	269348	Thalassiosira pseudonana	Metabolic Process Phosphate	C	29982	4.8	5.9
Molecular chaperone DnaK2	gi	33862260	Prochlorococcus marinus	Folding Protein	S	68202	4.8	2.5
Nitrate reductase	jgi	25299	Thalassiosira pseudonana	Transport Electron	S	101408	5.9	1

Oxygen-evolving enhancer protein 1 precursor	jgi	34830	Thalassiosira pseudonana	Photosynthesis	C	29136	5.2	8.4
Phosphoglycerate kinase precursor	jgi	35712	Thalassiosira pseudonana	Glycolysis	S	42256	5.0	4.5
Phosphoglycerate mutase 1	jgi	27850	Thalassiosira pseudonana	Glycolysis	S	32465	6.1	3.1
Phosphoribosyl-pyrophosphate synthetase	jgi	26109	Thalassiosira pseudonana	Biosynthesis	S	33425	8.4	3.2
Photosystem I protein PsaD	gi	33862134	Prochlorococcus marinus	Photosynthesis	C	15665	6.4	8.6
Photosystem II 10 kDa phosphoprotein	gi	118411116	Thalassiosira pseudonana	Photosynthesis	C	7388	6.0	21.2
Photosystem II chlorophyll A core antenna apoprotein	gi	118411113	Thalassiosira pseudonana	Photosynthesis	C	56408	6.5	2
Photosystem II chlorophyll A core antenna apoprotein CP43	gi	118411149	Thalassiosira pseudonana	Photosynthesis	C	51845	7.7	8.7
Photosystem II reaction center protein D1	gi	118411180	Thalassiosira pseudonana	Photosystem	C	39699	5.3	6.7
Photosystem II reaction center protein D2	gi	118411148	Thalassiosira pseudonana	Photosystem	C	39064	5.6	3.7
Photosystem II stability/assembly factor HCF136	jgi	38769	Thalassiosira pseudonana	Photosystem	C	40327	5.2	3.8
Predicted translation elongation factor G	jgi	25629	Thalassiosira pseudonana	Translation	S	86389	5.0	3.5
Protein product unnamed	jgi	29007	Thalassiosira pseudonana	Translation	M	18378	10.5	5.6
Putative ribosomal protein L12e	jgi	39424	Thalassiosira pseudonana	Translation	R	17412	9.1	9.1
Putative S-adenosyl-L-homocysteinas protein	jgi	28496	Thalassiosira pseudonana	Metabolic Process	S	52309	5.1	2.5
Pyruvate kinase	jgi	22345	Thalassiosira pseudonana	Glycolysis	C	57892	5.2	3.3
Ribosomal protein S3	jgi	28049	Thalassiosira pseudonana	Binding RNA	C	29161	9.3	13.3
Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	gi	118411104	Thalassiosira pseudonana	Photosynthesis	C	54325	6.2	17.3
Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	gi	118411103	Thalassiosira pseudonana	Photosynthesis	C	15843	5.1	10.1
RsuA	jgi	269764	Thalassiosira pseudonana	Binding RNA	S	36568	6.2	4
S-adenosyl methionine synthetase	jgi	21815	Thalassiosira pseudonana	Transport	S	50359	5.2	3.2
Transaldolase	jgi	27187	Thalassiosira pseudonana	Metabolic Process Carbohydrate	S	34855	4.8	4.1
Translation elongation factor 1 alpha	jgi	3858	Thalassiosira pseudonana	Translation	S	47815	7.2	2.5
Translation elongation factor Tu	gi	118411218	Thalassiosira pseudonana	Translation	C	44458	4.9	6.1
Triosephosphate isomerase/glyceraldehyde-3-phosphate dehydrogenase precursor	jgi	28239	Thalassiosira pseudonana	Metabolic Process	S	65308	5.6	2.3
Tubulin alpha-2 chain	jgi	29304	Thalassiosira pseudonana	Structure	S	49904	5.0	2.2
Tubulin beta chain	jgi	8069	Thalassiosira pseudonana	Structure	S	49497	4.9	5.6
Ubiquitin	jgi	40669	Thalassiosira pseudonana	Modification Protein	S	17567	9.9	8.5
Vacuolar ATP synthase 16 kDa proteolipid subunit	jgi	2233	Thalassiosira pseudonana	Transport Proton	S	16720	5.6	10.8
Vacuolar ATP synthase	jgi	37123	Thalassiosira	Transport	S	68343	5.0	2.4

subunit A			pseudonana	Proton				
Vacuolar ATPase B subunit	jgi	40522	Thalassiosira pseudonana	Transport Proton	S	56064	5.9	3
Vacuolar proton translocating ATPase A subunit, putative	jgi	40728	Thalassiosira pseudonana	Transport Proton	S	92414	5.3	1.6
Vacuolar-type H+-pyrophosphatase	jgi	32586	Thalassiosira pseudonana	Transport Proton	S	67969	4.6	1.7

Post Bloom Shelf Sediment

Protein	Annotation		Species	Function	Comp	MW	pI	SC
3-deoxy-7-phosphoheptulonate synthase	jgi	2790	Thalassiosira pseudonana	Transport Electron	S	53939	6.0	2.5
6-phosphogluconate dehydrogenase	jgi	33343	Thalassiosira pseudonana	Dehydrogenase	S	53348	5.6	4.7
Actin A	jgi	25772	Thalassiosira pseudonana	Binding Protein	S	41791	5.0	9.0
Adenosinetriphosphatase	jgi	40156	Thalassiosira pseudonana	Transport Proton	S	39935	7.6	7.4
ATP synthase CF0 B' chain subunit II	gi	118411109	Thalassiosira pseudonana	Transport Proton	C	17373	4.6	9.0
ATP synthase CF1 alpha chain	gi	118411112	Thalassiosira pseudonana	Transport Ion	C	53989	5.0	10.1
ATP synthase CF1 beta chain	gi	118411134	Thalassiosira pseudonana	Transport Ion	C	51143	4.7	20.5
ATP/ADP translocator	jgi	39143	Thalassiosira pseudonana	Transport	M	32254	9.4	3.6
Cell wall surface anchor family protein	jgi	6962	Thalassiosira pseudonana	Structure	S	75600	9.7	4.9
CG11154-PA, isoform A	jgi	41256	Thalassiosira pseudonana	Transport Proton	U	53388	5.1	8.0
Chloroplast light harvesting protein isoform 15	jgi	2845	Thalassiosira pseudonana	Photosynthesis	C	21873	5.1	5.4
Cytochrome b6	gi	118411154	Thalassiosira pseudonana	Transport Electron	C	23906	9.2	6.0
DNA binding	jgi	29950	Thalassiosira pseudonana	Binding DNA	C	15300	11.3	13.2
DNA-directed RNA polymerase beta prime chain	gi	71083809	Candidatus Pelagibacter ubique	Translation	S	154411	8.8	1.4
Domain specific binding protein 14-3-3	jgi	26146	Thalassiosira pseudonana	Binding Protein Domain Specific	S	27869	4.6	7.3
Eukaryotic translation initiation factor 4A2 isoform 2	jgi	9716	Thalassiosira pseudonana	Binding Nucleic Acid	S	42405	5.6	2.7
F0F1 ATP synthase subunit alpha	gi	33862007	Prochlorococcus marinus	Transport Proton	S	54306	4.9	4.6
F0F1 ATP synthase subunit beta	gi	71082935	Candidatus Pelagibacter ubique	Transport Proton	S	50696	4.9	10.0
Fucoxanthin chlorophyl a/c protein	jgi	38667	Thalassiosira pseudonana	Photosynthesis	C	21807	4.8	13.2
Fucoxanthin chlorophyll a /c protein	jgi	42962	Thalassiosira pseudonana	Photosynthesis	C	21515	5.1	6.5
Fucoxanthin chlorophyll a/c binding protein	jgi	268127	Thalassiosira pseudonana	Photosynthesis	C	22628	4.8	12.9
Glyceraldehyde-3-phosphate dehydrogenase precursor	jgi	31383	Thalassiosira pseudonana	Glycolysis	S	39587	5.3	9.1
Heat shock protein Hsp90	jgi	6285	Thalassiosira pseudonana	Folding Protein	S	80242	4.7	4.0
Histone 3	jgi	3183	Thalassiosira	Binding DNA	C	15312	11.4	13.2

			pseudonana					
Histone H2A.1	jgi	19793	Thalassiosira pseudonana	Binding DNA	N	13053	10.4	18.5
Histone H4	jgi	3184	Thalassiosira pseudonana	Binding DNA	N	11384	11.5	42.7
Hypothetical protein CBG01077	jgi	22792	Thalassiosira pseudonana	Transport	S	22486	6.7	10.9
Hypothetical protein CBG08717	jgi	269322	Thalassiosira pseudonana	Transport Proton	U	58009	5.8	9.2
Hypothetical protein DEHA0F19712g	jgi	27352	Thalassiosira pseudonana	Binding DNA	N	13064	10.3	12.5
Hypothetical protein FG01081.1	jgi	25949	Thalassiosira pseudonana	Translation	U	20124	9.8	7.9
Manganese superoxide dismutase	jgi	32874	Thalassiosira pseudonana	Metabolic Process	M	27061	5.5	5.3
Photosystem I p700 chlorophyll A apoprotein A	gi	118411096	Thalassiosira pseudonana	Photosynthesis	C	83642	7.3	3.5
Photosystem I p700 chlorophyll A apoprotein B	gi	118411097	Thalassiosira pseudonana	Photosynthesis	C	82090	7.6	4.5
Photosystem I protein F	gi	118411168	Thalassiosira pseudonana	Photosynthesis	C	20362	8.9	8.1
Photosystem I protein L	gi	118411163	Thalassiosira pseudonana	Photosynthesis	C	15704	9.3	12.2
Photosystem II 10 kDa phosphoprotein	gi	118411116	Thalassiosira pseudonana	Photosynthesis	C	7388	6.0	21.2
Photosystem II chlorophyll A core antenna apoprotein	gi	118411113	Thalassiosira pseudonana	Photosynthesis	C	56408	6.5	18.9
Photosystem II chlorophyll A core antenna apoprotein CP43	gi	118411149	Thalassiosira pseudonana	Photosynthesis	C	51845	7.7	15.9
Photosystem II reaction center protein D1	gi	118411180	Thalassiosira pseudonana	Photosystem	C	39699	5.3	6.4
Photosystem II reaction center protein D2	gi	118411148	Thalassiosira pseudonana	Photosystem	C	39064	5.6	6.3
Rab family GTPase Rab8	jgi	33126	Thalassiosira pseudonana	Transport	S	20382	7.7	12.2
RAB small monomeric GTPase	jgi	35818	Thalassiosira pseudonana	Transport	S	20626	6.6	11.1
Ribosomal protein S18	jgi	26893	Thalassiosira pseudonana	Structure Ribosome	R	17159	10.8	13.7
Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	gi	118411104	Thalassiosira pseudonana	Photosynthesis	C	54325	6.2	25.3
S-adenosyl methionine synthetase	jgi	21815	Thalassiosira pseudonana	Transport	S	50359	5.2	3.2
Translation elongation factor 1 alpha	jgi	3858	Thalassiosira pseudonana	Translation	S	47815	7.2	2.5
Tubulin alpha-2 chain	jgi	29304	Thalassiosira pseudonana	Structure	S	49904	5.0	2.9
Tubulin beta chain	jgi	8069	Thalassiosira pseudonana	Structure	S	49497	4.9	11.5
Ubiquitin	jgi	40669	Thalassiosira pseudonana	Modification Protein	S	17567	9.9	22.2
Unknown	jgi	39299	Thalassiosira pseudonana	Binding GTP	U	20905	6.8	25.1
Vacuolar ATP synthase 16 kDa proteolipid subunit	jgi	2233	Thalassiosira pseudonana	Transport Proton	S	16720	5.6	10.8
Vacuolar proton-inorganic pyrophosphatase	jgi	39520	Thalassiosira pseudonana	Transport Proton	S	70120	5.0	2.4

Post Bloom Basin Sediment

Protein	Annotation		Species	Function	Comp	MW	pI	SC
ATP synthase CF1 beta chain	gi	118411134	Thalassiosira pseudonana	Transport Ion	C	51143	4.7	5.3
DNA-directed RNA polymerase subunit gamma	gi	33862040	Prochloro-coccus marinus	Translation	S	72335	6.6	1.7
Fucoxanthin chlorophyl a/c protein	jgi	38667	Thalassiosira pseudonana	Photosynthesis	C	21807	4.8	6.8
Fucoxanthin chlorophyll a /c protein	jgi	38494	Thalassiosira pseudonana	Photosynthesis	C	20354	4.5	6.8
Fucoxanthin chlorophyll a/c binding protein	jgi	268127	Thalassiosira pseudonana	Photosynthesis	C	22628	4.8	6.2
Histone H2A.1	jgi	19793	Thalassiosira pseudonana	DNA Binding	N	13053	10.4	7.3
Histone H4	jgi	3184	Thalassiosira pseudonana	DNA Binding	N	11384	11.5	41.7
Hypothetical protein DDB0187116	jgi	25297	Thalassiosira pseudonana	Metabolic Process Lipid	U	369027	6.3	0.9
Intracellular membrane-associated calcium-independent phospholipase A2 gamma	jgi	23984	Thalassiosira pseudonana	Metabolic Process Lipid	S	92973	5.4	2.3
OmpA family protein	gi	71083303	Candidatus Pelagibacter ubique	Membrane	S	17499	9.6	5.7
Oxygen-evolving enhancer protein 1 precursor	jgi	34830	Thalassiosira pseudonana	Photosynthesis	S	29136	5.2	7.6
Photosystem I p700 chlorophyll A apoprotein A	gi	118411096	Thalassiosira pseudonana	Photosynthesis	C	83642	7.3	1.9
Photosystem I p700 chlorophyll A apoprotein B	gi	118411097	Thalassiosira pseudonana	Photosynthesis	C	82090	7.6	3.4
Photosystem II 10 kDa phosphoprotein	gi	118411116	Thalassiosira pseudonana	Photosynthesis	C	7388	6.0	21.2
Photosystem II chlorophyll A core antenna apoprotein	gi	118411113	Thalassiosira pseudonana	Photosynthesis	C	56408	6.5	18.7
Photosystem II chlorophyll A core antenna apoprotein CP43	gi	118411149	Thalassiosira pseudonana	Photosynthesis	C	51845	7.7	9.6
Photosystem II reaction center protein D1	gi	118411180	Thalassiosira pseudonana	Photosystem	C	39699	5.3	3.1
Photosystem II reaction center protein D2	gi	118411148	Thalassiosira pseudonana	Photosystem	C	39064	5.6	13.7
TFIID subunit	jgi	3021	Thalassiosira pseudonana	Transcription	N	55578	6.6	4.4
Transcription termination factor Rho	gi	71083054	Candidatus Pelagibacter ubique	Transcription	S	47060	8.4	2.6
Translation elongation factor 1 alpha	jgi	3858	Thalassiosira pseudonana	Translation	S	47815	7.2	2.5
Tubulin beta chain	jgi	8069	Thalassiosira pseudonana	Structure	U	49497	4.9	2.7
Ubiquitin	jgi	40669	Thalassiosira pseudonana	Modification Protein	S	17567	9.9	11.8
Vacuolar ATP synthase 16 kDa proteolipid subunit	jgi	2233	Thalassiosira pseudonana	Transport Proton	S	16720	5.6	10.8

Over Wintered Shelf Sediment

Protein	Annotation		Species	Function	Comp	MW	pI	SC
Actin A	jgi	25772	Thalassiosira pseudonana	Binding Protein	S	41791	5.0	7.2
ATP synthase CF1 beta chain	gi	118411134	Thalassiosira pseudonana	Transport Ion	C	51143	4.7	8.0
DNA-directed RNA polymerase subunit gamma	gi	33862040	Prochloro-coccus marinus	Translation	S	72335	6.6	1.7
Fucoxanthin chlorophyl a/c protein	jgi	38667	Thalassiosira pseudonana	Photosynthesis	C	21807	4.8	6.8
Fucoxanthin chlorophyll a /c protein	jgi	42962	Thalassiosira pseudonana	Photosynthesis	C	21515	5.1	6.5
Fucoxanthin chlorophyll a/c binding protein	jgi	268127	Thalassiosira pseudonana	Photosynthesis	C	22628	4.8	6.2
Heat shock protein 70	jgi	269120	Thalassiosira pseudonana	Folding Protein	S	71187	4.8	4.1
Histone H2A.1	jgi	19793	Thalassiosira pseudonana	Binding DNA	N	13053	10.4	7.3
Histone H4	jgi	3184	Thalassiosira pseudonana	Binding DNA	N	11384	11.5	40.8
Hypothetical protein UM00510.1	jgi	261141	Thalassiosira pseudonana	Binging DNA	S	29752	8.9	4.4
Photosystem I p700 chlorophyll A apoprotein A	gi	118411096	Thalassiosira pseudonana	Photosynthesis	C	83642	7.3	1.9
Photosystem I p700 chlorophyll A apoprotein B	gi	118411097	Thalassiosira pseudonana	Photosynthesis	C	82090	7.6	2.9
Photosystem I protein L	gi	118411163	Thalassiosira pseudonana	Photosynthesis	C	15704	9.3	12.2
Photosystem II 10 kDa phosphoprotein	gi	118411116	Thalassiosira pseudonana	Photosynthesis	C	7388	6.0	21.2
Photosystem II chlorophyll A core antenna apoprotein	gi	118411113	Thalassiosira pseudonana	Photosynthesis	C	56408	6.5	5.9
Photosystem II chlorophyll A core antenna apoprotein CP43	gi	118411149	Thalassiosira pseudonana	Photosynthesis	C	51845	7.7	6.4
Photosystem II reaction center protein D1	gi	118411180	Thalassiosira pseudonana	Photosystem	C	39699	5.3	3.1
Photosystem II reaction center protein D2	gi	118411148	Thalassiosira pseudonana	Photosystem	C	39064	5.6	6.3
Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	gi	118411104	Thalassiosira pseudonana	Photosynthesis	C	54325	6.2	11.0
Translation elongation factor 1 alpha	jgi	3858	Thalassiosira pseudonana	Translation	S	47815	7.2	2.5
Tubulin beta chain	jgi	8069	Thalassiosira pseudonana	Structure	U	49497	4.9	5.4
Ubiquitin	jgi	40669	Thalassiosira pseudonana	Modification Protein	S	17567	9.9	15.7
Vacuolar ATP synthase 16 kDa proteolipid subunit	jgi	2233	Thalassiosira pseudonana	Transport Proton	S	16720	5.6	10.8

APPENDIX 4-2

(A) Total hydrolyzable amino acid mole percent distribution of suspended particle, sediment trap, and sediment samples; (B) Tabulated amino acid mole percent distribution of identified proteins; (C) Tabulated amino acid mole percent distribution of identified protein transmembrane regions.

(A)

	Chl Max POC	50m POC	100m POC	40m Trap	60m Trap	100m Trap	PBS	PBB	OWS
Ala	11.11	11.80	12.21	9.11	10.01	9.66	7.67	7.57	7.20
Gly	7.82	13.12	16.11	9.52	10.02	11.37	17.06	16.93	17.17
Val	4.62	4.47	4.01	5.93	6.18	5.65	4.71	4.94	5.18
Leu	6.25	6.41	4.43	7.31	7.66	6.46	3.39	3.44	3.12
Ile	4.04	2.57	1.61	4.53	4.50	3.88	3.68	4.06	3.76
Thr	3.80	2.06	2.77	3.96	3.79	4.03	5.83	5.13	5.43
Pro	5.12	6.07	5.75	4.83	5.19	5.28	4.42	4.78	4.43
Asp/Asn	8.73	6.82	6.50	8.30	7.12	8.50	10.05	10.07	10.30
Phe	3.09	1.77	0.88	3.73	3.13	2.46	1.92	1.89	1.92
Glu/Gln	12.74	4.99	3.75	7.80	5.60	6.49	6.48	5.68	6.94
Lys	4.38	1.80	4.15	3.33	2.53	2.07	2.04	1.94	2.35
Tyr	2.80	4.73	8.68	3.85	4.24	4.16	2.86	0.87	3.00

(B)

	Chl Max POC	50m POC	100m POC	40m Trap	60m Trap	100m Trap	PBS	PBB	OWS
Leu	8.20	8.12	8.28	8.29	7.93	8.12	8.86	9.40	9.22
Gly	8.46	8.65	9.48	8.57	9.02	8.64	8.87	8.60	9.50
Ala	9.26	8.93	10.12	9.22	9.51	8.93	9.32	8.10	8.72
Phe	3.72	4.29	4.73	3.88	3.79	3.93	4.39	4.88	5.38
Ile	6.14	6.19	6.16	6.30	6.32	6.37	6.55	5.96	6.65
Val	7.46	7.36	7.27	7.43	7.69	7.52	6.89	7.00	6.56
Ser	6.31	7.27	5.94	6.06	6.00	6.05	6.52	7.45	5.92
His	1.62	1.82	1.49	1.66	1.43	1.71	1.91	2.49	2.50
Thr	5.76	6.15	6.25	5.77	5.80	5.96	5.73	5.65	5.57
Met	2.78	2.65	2.94	2.84	2.91	2.79	2.71	2.73	2.93
Trp	0.87	1.20	1.12	0.90	0.91	1.01	1.28	1.57	1.93
Tyr	2.80	2.90	3.04	2.81	2.82	2.90	2.94	2.67	2.99
Pro	4.25	4.43	3.95	4.10	4.09	4.22	3.93	4.25	4.48
Cys	1.31	1.29	1.04	1.23	1.22	1.24	1.13	0.97	1.00
Asn	3.97	3.66	3.54	3.85	3.73	3.84	3.56	4.01	4.00
Gln	3.23	3.19	3.62	3.28	3.13	3.18	3.42	3.19	3.30
Arg	4.82	4.81	4.57	4.76	4.68	4.83	4.83	5.24	4.60
Glu	6.93	6.16	5.58	7.05	6.94	6.74	6.11	5.86	5.52

Lys	6.06	5.23	5.38	6.04	5.66	6.13	5.70	4.78	4.47
Asp	6.05	5.72	5.50	5.96	6.42	5.87	5.34	5.18	4.77

(C)

	Chl Max POC	50m POC	100m POC	40m Trap	60m Trap	100m Trap	PBS	PBB	OWS
Leu (L)	13.80	14.38	14.33	14.61	14.16	14.16	14.37	14.30	14.56
Gly (G)	12.49	13.05	12.54	12.94	13.48	12.09	12.46	12.33	12.63
Ala (A)	13.42	13.45	15.64	12.71	11.24	12.42	12.06	13.05	12.29
Phe (F)	9.63	9.32	8.79	9.78	10.11	9.62	9.46	10.01	10.02
Ile (I)	10.50	9.72	9.28	10.64	8.76	9.82	9.92	9.56	9.85
Val (V)	9.70	8.92	9.45	8.17	8.99	10.42	8.94	7.77	7.58
Ser (S)	6.28	7.59	7.82	5.92	6.07	5.95	5.89	5.45	5.56
His (H)	1.68	2.66	1.95	3.22	4.27	2.07	4.15	5.90	5.64
Thr (T)	5.10	3.46	4.40	3.68	3.60	4.74	4.21	3.66	3.79
Met (M)	2.80	4.13	4.56	4.08	3.37	4.14	3.75	3.75	3.70
Trp (W)	2.18	2.93	2.61	2.82	4.49	2.40	3.17	4.02	3.87
Tyr (Y)	3.60	3.20	2.28	3.45	3.82	3.54	3.75	3.40	3.54
Pro (P)	2.86	2.26	2.28	2.59	3.15	2.54	2.37	2.06	2.10
Cys (C)	1.93	1.73	1.14	1.61	1.80	1.94	1.62	1.52	1.60
Asn (N)	1.18	0.80	0.81	1.27	0.90	1.14	1.04	1.07	1.01
Gln (Q)	0.87	0.53	0.65	0.92	0.22	1.07	0.75	0.63	0.67
Arg (R)	0.62	0.40	0.33	0.52	0.45	0.67	0.81	0.54	0.59
Glu (E)	0.62	0.53	0.49	0.46	0.45	0.40	0.63	0.36	0.42
Lys (K)	0.37	0.53	0.16	0.40	0.45	0.40	0.40	0.54	0.42
Asp (D)	0.37	0.40	0.49	0.23	0.22	0.47	0.23	0.09	0.17

APPENDIX 5-1

Amino acid distribution (mole %) of phytoplankton material from each time point of the 11 day and 53 day incubations, standard error in ( ).

11 Day Incubation

Days	0	0.5	1	3	5	7	9	11
	14.92	9.8	9.63	8.55	8.52	9.48	10.08	9.6
Alanine	(0.33)	(0.18)	(0.12)	(0.46)	(0.17)	(0.06)	(0.77)	(0.16)
	7.37	7.83	8.35	8.42	8.3	8.96	8.86	9.48
Glycine	(0.09)	(0.05)	(0.13)	(0.13)	(0.19)	(0.12)	(0.17)	(0.11)
	6.93	5.42	5.37	5.84	5.69	5.83	5.55	5.74
Valine	(0.33)	(0.11)	(0.08)	(0.15)	(0.08)	(0.10)	(0.08)	(0.12)
	6.56	6.32	6.59	6.64	6.48	6.83	6.69	6.62
Leucine	(0.17)	(0.05)	(0.01)	(0.09)	(0.10)	(0.04)	(0.01)	(0.09)
	4.35	4.6	4.49	4.98	4.65	4.35	4.11	4.3
Isoleucine	(0.16)	(0.05)	(0.30)	(0.10)	(0.21)	(0.06)	(0.16)	(0.03)
	3.66	7.87	7.29	7.43	8.35	5.93	5.59	6.06
Threonine	(0.25)	(0.54)	(0.30)	(0.35)	(0.42)	(0.14)	(0.32)	(0.04)
	4.42	4.17	4.22	3.81	3.58	4.34	4.21	4.16
Proline	(0.09)	(0.03)	(0.02)	(0.04)	(0.10)	(0.12)	(0.03)	(0.07)
Aspartic	5.56	9.53	9.53	10.25	10.07	9.45	9.37	9.37
Acid/Asparagine	(0.24)	(0.13)	(0.10)	(0.30)	(0.25)	(0.05)	(0.28)	(0.15)
	2.33	3.13	3.32	3.69	3.49	3.27	3.26	3.25
Phenylalanine	(0.04)	(0.06)	(0.01)	(0.13)	(0.08)	(0.06)	(0.02)	(0.11)
Glutamic	10.18	12.32	11	10.47	12.22	10.21	11.01	10.31
Acid/Glutamine	(1.40)	(1.01)	(0.16)	(0.17)	(0.59)	(0.87)	(1.22)	(0.97)
	3.94	3.18	3.21	3.45	4.53	2.94	3.46	2.81
Lysine	(0.72)	(0.12)	(0.11)	(0.08)	(0.31)	(0.38)	(0.27)	(0.47)
	3.65	1.99	1.22	1.16	2.28	0.68	0.91	0.7
Tyrosine	(1.92)	(0.25)	(0.03)	(0.13)	(0.14)	(0.18)	(0.13)	(0.07)

53 Day Incubation

Days	0	5	12	22	35	47	53
	17.93	15.3	15.75	12.87	15.43	15.45	15.3
Alanine	(1.79)	(0.51)	(1.52)	(0.70)	(0.27)	(0.16)	(0.72)
	13.57	13.24	15.99	11.62	15.37	16.9	17.14
Glycine	(0.16)	(0.55)	(2.47)	(0.51)	(0.72)	(0.29)	(0.63)
	6.18	6.91	5.96	7.72	10.68	11.66	11.26
Valine	(0.44)	(0.39)	(0.63)	(0.25)	(0.68)	(0.03)	(0.76)
	9.12	9.26	9.24	7.81	9.21	9.8	9.26
Leucine	(0.35)	(0.60)	(0.92)	(0.32)	(0.44)	(0.18)	(0.58)
	4.37	3.56	4.23	4.28	3.89	3.72	3.53
Isoleucine	(0.13)	(0.23)	(0.19)	(0.12)	(0.03)	(0.16)	(0.22)
	3.42	4.18	4.75	7.21	6.7	6.79	6.99
Threonine	(0.44)	(0.49)	(0.67)	(0.39)	(0.74)	(0.48)	(1.44)
	6.04	5.99	7.06	4.51	3.02	2.62	3.13
Proline	(0.62)	(0.20)	(1.57)	(0.08)	(0.49)	(0.27)	(0.55)
Aspartic	10.18	10.48	11.39	10.18	4.97	4.88	7.41

Acid/Asparagine	(0.88)	(0.46)	(0.95)	(0.59)	(0.98)	(0.86)	(2.08)
	3.69	3.46	3.96	4.16	5.31	5.32	4.98
Phenylalanine	(0.41)	(0.28)	(0.12)	(0.10)	(0.21)	(0.33)	(0.13)
Glutamic	12.74	14.34	13.63	19.57	12.16	10.61	10.11
Acid/Glutamine	(0.57)	(0.80)	(3.72)	(1.41)	(0.96)	(0.77)	(0.28)
	5.43	5.88	4.18	5.11	5.48	4.54	3.63
Lysine	(0.57)	(0.44)	(1.18)	(0.18)	(0.71)	(0.65)	(0.64)
	7.47	7.41	3.87	4.99	7.78	7.69	7.27
Tyrosine	(0.35)	(3.37)	(1.54)	(0.38)	(0.56)	(1.52)	(0.95)

References

Abramson L., Lee, C., Liu, Z., Wakeham, S. G., Szlosek, J. (2010) Exchange between suspended and sinking particles in the northwest Mediterranean as inferred from the organic composition of in situ pump and sediment trap samples. *Limnology and Oceanography* **55**(2), 725-739.

Aebersold R., Goodlett, D. R. (2001) Mass spectrometry in proteomics. *Chemical Reviews* **101**(2), 269-295.

Aebersold R., Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**(6928), 198-207.

Ahmed N., Thornalley, P. J. (2003) Advanced glycation endproducts: what is their relevance to diabetic complications? *Diabetes, Obesity, and Metabolism* **9**, 233-245.

Armbrust E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S. G., Allen, A. E., Apt, K. E., Bechner, M., Brzezinski, M. A., Chaal, B. K., Chiovitti, A., Davis, A. K., Demarest, M. S., Detter, J. C., Glavina, T., Goodstein, D., Hadi, M. Z., Hellsten, U., Hildebrand, M., Jenkins, B. D., Jurka, J., Kapitonov, V. V., Kroger, N., Lau, W. W. Y., Lane, T. W., Larimer, F. W., Lippmeier, J. C., Lucas, S., Medina, M., Montsant, A., Obornik, M., Parker, M. S., Palenik, B., Pazour, G. J., Richardson, P. M., Rynearson, T. A., Saito, M. A., Schwartz, D. C., Thamatrakoln, K., Valentin, K., Vardi, A., Wilkerson, F. P., Rokhsar, D. S. (2004) The genome of the diatom Thalassiosira pseudonana: Ecology, evolution, and metabolism. *Science* **306**(5693), 79-86.

Ashburner M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., Sherlock, G. (2000) Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**(1), 25-29.

Banahan S., Goerging, J. (1986) The production of biogenic silica and its accumulation on the southeastern Bering Sea shelf. *Continental Shelf Research* **5**, 199-213.

Bandeira N., Tsur, D., Frank, A., Pevzner, P. A. (2007) Protein identification by spectral networks analysis. *Proceedings of the National Academy of Sciences* **104**(15), 6140-6145.

Bastida F., Moreno, J. L., Nicolas, C., Hernandez, T., Carcia, C. (2009) Soil metaproteomics: a review of an emerging environmental science. Significance, methodology and perspectives. *European Journal of Soil Science* **60**, 845-859.

Bause E., Hettkamp, H. (1979) Primary structural requirements for *N*-glycosylation of peptides in rat-liver. *FEBS Letters* **108**(2), 341-344.

Beaufort L., Probert, I., de Garidel-Thoron, T., Bendif, E. M., Ruiz-Pino, D., Metzl, N., Goyet, C., Buchet, N., Coupel, P., Grelaud, M., Rost, B., Rickaby, R. E. M., de Vargas, C. (2011) Sensitivity of coccolithophores to carbonate chemistry and ocean acidification. *Nature* **476**(7358), 80-83.

Bechtold U., Rabbani, N., Mullineaux, P. M., Thornalley, P. J. (2009) Quantitative measurement of specific biomarkers for protein oxidation, nitration and glycation in Arabidopsis leaves. *The Plant Journal* **59**, 661-671.

Belluomini G., Branca, M., Calderoni, G., Schnitzer, M. (1986) Distribution and geochemical significance of amino-acids and amino-sugars in a clay suite of the pliocene pleistocene age from central italy. *Organic Geochemistry* **9**(3), 127-133.

Benndorf D., Balcke, G. U., Harms, H., Bergen, M. (2007) Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. *ISME Journal* **1**, 224-234.

Benner R., Pakulski, J. D., McCarthy, M., Hedges, J. I., Hatcher, P. G. (1992) Bulk Chemical Characteristics of Dissolved Organic matter in the Ocean. *Science* **255**, 1561-1564.

Bern M., Goldberg, D., McDonald, W. H., Yates, J. R. III. (2004) *Bioinformatics* **20**, 149.

Birktoft J. J., Breddam, K. (1994) Proteolytic Enzymes: Serine and Cysteine Peptidases. In *Methods in Enzymology*, Vol. 244 (ed. A. J. Barrett), pp. pp. 114-126. Academic Press.

Borg I., Groenen, P. (2005) Modern Multidimensional Scaling: theory and applications, pp. 207-212. Springer-Verlag.

Boyd S. R. (2001) Nitrogen in future biosphere studies. *Chemical Geology* **176**, 1-30.

Brooks D. J., Fresco, J. R., Lesk, A. M., Singh, M. (2002) Evolution of amino acid frequencies in proteins over deep time: inferred order of introduction of amino acids into genetic code. *Molecular Biology and Evolution* **19**(10), 1645-1655.

Brown M. R. (1991) The amino-acid and sugar composition of 16 species of microalgae used in mariculture. *Journal of Experimental Marine Biology and Ecology* **145**(1), 79-99.

Burdige D. J., Martens, C. S. (1988) Biogeochemical cycling in an organic-rich coastal marine basin .10. the role of amino-acids in sedimentary carbon and nitrogen cycling. *Geochimica et Cosmochimica Acta* **52**(6), 1571-1584.

Burdige D. J. (2007) Preservation of organic matter in marine sediments: Controls, mechanisms, and an imbalance in sediment organic carbon budgets? *Chemical Reviews* **107**(2), 467-485.

Cervantes-Laurean D., Jacobson, E. L., Jacobson, M. K. (1996) Glycation and glycoxidation of histones by ADP-ribose. *Journal of Biological Chemistry* **271**(18), 10461-10469.

Chen J., Ryu, S., Gharib, S. A., Goodlett, D. R., Schnapp, L. M. (2008) Exploration of the normal human bronchioalveolar lavage fluid proteome. *Proteomics Clinical Applications* **2**(4), 585-595.

Chen M., Huang, Y. P., Cai, P. G., Guo, L. D. (2003) Particulate organic carbon export fluxes in the Canada Basin and Bering Sea as derived from $^{234}Th/^{238}U$ disequilibria. *Arctic* **56**(1), 32-44.

Cheng C. N., Shufeldt, R. C., Stevenson, F. J. (1975) Amino acid analysis of soils and sediments: extraction and desalting. *Soil Biology and Biochemistry* **7**(2), 143-151.

Chu F. K. (1986) Requirements of cleavage of high mannose oligosaccharides in glycoproteins by Peptide *N*-Glycosidase F. *Biological Chemistry* **261**(1), 172-177.

Collins M. J., Westbroek, P., Muyzer, G., Deleeuw, J. W. (1992) Experimental-evidence for condensation-reactions between sugars and proteins in carbonate skeletons. *Geochimica et Cosmochimica Acta* **56**(4), 1539-1544.

Collins M. J., Bishop, A. N., Farrimond, P. (1995) Sorption by mineral surfaces: Rebirth of the classical condensation pathway for kerogen formation? *Geochimica et Cosmochimica Acta* **59**(11), 2387-2391.

Consortium G. O. (2008) The Gene Ontology project in 2008. *Nucleic Acids Research* **36**, D440-D444.

Coughenour H. D., Spaulding, R. S., Thompson, C. M. (2004) The synaptic vesicle proteome: A comparative study in membrane protein identification. *Proteomics* **4**(10), 3141-3155.

Cowie G. L., and Hedges J. I. (1992a) Sources and Reactivities of Amino Acids in a Coastal Marine Environment. *Limnology and Oceanography* **37**, 703-724.

Cowie G. L., Hedges, J. I. (1992b) Improved amino acid quantification in environmental-samples - charge-matched recovery standards and reduced analysis time. *Marine Chemistry* **37**(3-4), 223-238.

Cowie G. L., Hedges, J. I. (1994) Biochemical indicators of diagenetic alteration in natural organic matter mixtures. *Nature* **369**, 304.

Craig O. E., Collins, M. J. (2000) An improved method for the immunological detection of mineral bound protein using hydrofluoric acid and direct capture. *Journal of Immunological Methods* **236**, 89-97.

Criquet S., Farnet, A. M., Ferre, E. (2002) Protein measurement in forest litter. *Biology & Fertility of Soils* **35**, 307-313.

Cronin J. R., Morris, R. J. (1981) Rapid formation of humic material from diatom debris. In *Coastal upwelling, its sediment record* (ed. E. T. Suess, J.), pp. 485-496. Plenum.

Dauwe B., Middelburg, J. T. (1998) Amino Acids and Hexosamines as Indicators of Organic Matter Degradation State in North Sea Sediments. *Limnology and Oceanography* **43**(5), 782-798.

Davis J. A. (1982) Adsorption of natural dissolved organic matter at the oxider/water interface. *Geochimica et Cosmochimica Acta* **46**, 2381-2393.

Davis M. T., Spahr, C. S., McGinley, M. D., Robinson, J. H., Bures, E. J., Beierle, J., Mort, J., Yu, W., Luethy, R., Patterson, S. D. (2001) Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry - II. Limitations of complex mixture analyses. *Proteomics* **1**(1), 108-117.

de Leeuw J. W., Largeau, C. (1993) A review of macromolecular organic compounds that comprise living organisms and their role in kerogen, coal and petroleum formation. In *Organic Geochemistry* (ed. M. H. Engel, Macko, S. A.), pp. 23-72. Plenum.

Deming J. W., Colwell, R. R. (1982) Barophillic bacterial associated with digestive tracts of abyssal holothurions. *Applied and Environmental Microbiology* **44**, 1222-1230.

Dong H. P., Wang, D. Z., Dai, M., Hong, H. S. (2010) Characterization of particulate organic matters in the water column of the South China Sea using a shotgun proteomic approach. *Limnology and Oceanography* **55**(4), 1565-1578.

Drapeau G. R., Boily, Y., Houmard, J. (1972) Purification and properties of an extracellular protease of Staphylococcus aureus. *Journal of Biological Chemistry* **247**, 6720-6726.

Dunne J. P., Armstrong, R.A., Gnanadesikan, A., Sarmiento, J. L. (2005) Empirical and mechanistic models for the particle export ratio. *Global Biogeochemical Cycles* **19**(4), GB4026.

Dunning J. D., Herren, B. J., Tipps, R. W., Snyder, R. S. (1982) Fractionation of mineral species by electrophoresis. *Journal of Geophysical Research* **87**, 781-788.

Emanuelsson O., Brunak, S., von Heijne, G., Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**, 953-971.

Eng J. K., McCormack, A. L., Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino-acid-sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **5**, 976-989.

Eng J. K., Fischer, B., Grossmann, J., MacCoss, M. J. (2008) A Fast SEQUEST Cross Correlation Algorithm. *Journal of Proteome Research* **7**(10), 4598-4602.

Fairbanks G., Steck, T. L., Wallach, D. F. H. (1971) Electrophoretic analysis of the major proteins of the human erythrocyte membrane. *Biochemistry* **10**, 2606-2617.

Falkowski P. G. (1997) Evolution of the nitrogen cycle and its influence of the biological sequestration of CO2 in the ocean. *Nature* **387**, 272-275.

Feng Y. Y., Hare C. E., Leblanc K., Rose J. M., Zhang Y. H., DiTullio G. R., Lee P. A., Wilhelm S. W., Rowe J. M., Sun J., Nemcek N., Gueguen C., Passow U., Benner I., Brown C., Hutchins D. A. (2009) Effects of increased pCO_2 and temperature on the North Atlantic spring bloom. I. The phytoplankton community and biogeochemical response. *Marine Ecology Progress Series* **388**, 13-25.

Fenn J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**(4926), 64-71.

Fogel M. L., Tuross, N. (1999) Transformation of plant biochemicals to geological macromolecules during early diagenesis. *Oecologia* **120**(3), 336-346.

Fujii K., Luo, Y., Sasahira, T., Denda, A., Ohmori, H., Kuniyasu, H. (2009) Co-treatment with deoxycholic acid and azoxymethane accelerates secretion of HMGB1 in IEC6 intestinal epithelial cells. *Cell Proliferation* **42**(5), 701-709.

Galloway J. N. e. a. (2004) Nitrogen Cycles: Past, Present, and Future. *Biogeochemistry* **70**, 153-226.

Granum E., Raven, J. A., Leegood, R. C. (2005) How do marine diatoms fix 10 billion tonnes of inorganic carbon per year? *Canadian Journal of Botany-Revue Canadienne de Botanique* **83**(7), 898-908.

Graves P. R., Haystead, T. A. J. (2002) Molecular biologist's guide to proteomics. *Microbiology & Molecular Biology Reviews* **66**(1), 39-63.

Grebmeier J. M., McRoy, C. P., Feder, H. M. (1988) Pelagic-benthic coupling on the shelf of the northern Bering and Chukchi Seas. I. Food supply source and benthic biomass. *Marine Ecology Progress Series* **48**, 57-67.

Greenland D. J. (1971) Interactions between humic and fulvic acids and clays. *Soil Science* **111**, 34-41.

Griffiths R. P., Hayasaka, S. S., McNamara, T. M., Morita, R. Y. (1978) Relative microbial activity and bacterial concentrations in water and sediment samples taken in the Beaufort Sea. *Canadian Journal of Microbiology* **24**, 1217-1226.

Grossman A. R., Bhaya, D., Apt, K. E., Kehoe, D. M. (1995) Light-harvesting complexes in oxygenic photosynthesis: diversity, control, and evolution. *Annual Review of Genetics* **29**, 231-288.

Grutters M., van Raaphorst W., Helder W. (2001) Total hydrolysable amino acid mineralisation in sediments across the northeastern Atlantic continental slope (Goban Spur). *Deep-Sea Research Part I-Oceanographic Research Papers* **48**(3), 811-832.

Guinotte J. M., Fabry, V. J. (2008) Ocean acidification and its potential effects on marine ecosystems. In *Year in Ecology and Conservation Biology 2008*, Vol. 1134 (ed. R. S. Ostfeld, Schlesinger, W. H.), pp. 320-342. Blackwell Publishing.

Hadley C. (2003) Righting the wrongs. *EMBO* **4**, 829.

Hare C. E., Leblanc K., DiTullio G. R., Kudela R. M., Zhang Y., Lee P. A., Riseman S., Hutchins D. A. (2007) Consequences of increased temperature and CO_2 for phytoplankton community structure in the Bering Sea. *Marine Ecology Progress Series* **352**, 9-16.

Harris S. A., Robinson, J. P., Juniper, B. E. (2002) Genetic clues to the origin of the apple. *Trends in Genetics* **18**, 426-430.

Harvey H. R., Richardson, M. D., Patton, J. S. (1984) Lipid composition and vertical distribution of bacteria in aerobic sediments in Venezuelan Basin. *Deep-Sea Research Part A, Oceanographic Research Papers* **31**(4), 403-413.

Harvey H. R., Tuttle, J. H., Bell, J. T. (1995) Kinetics of phytoplankton decay during simulated sedimentation: Changes in biochemical composition and microbial activity under oxic and anoxic conditions. *Geochimica et Cosmochimica Acta* **59**(16), 3367-3377.

Hedges J. (1991) Lignin, cutin, amino acids and carbohydrate analyses of marine particulate organic matter. In *Marine particles: Analysis and characterization* (ed. D. C. H. a. D. W. Spencer), pp. 129-137. American Geophysical Union.

Hedges J. I., Keil, R. G. (1999) Organic geochemical perspectives on estuarine processes: sorption reactions and consequences. *Marine Chemistry* **65**, 55-65.

Henrichs S. M., Sugai, S. F. (1993) Adsorption of amino-acids and glucose by sediments of Resurrection Bay, Alaska, USA - functional-group effects. *Geochimica et Cosmochimica Acta* **57**(4), 823-835.

Henzel W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., Watanabe, C. (1993) Identifying proteins from 2-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proceedings of the National Academy of Sciences of the United States of America* **90**(11), 5011-5015.

Highsmith R. C., and Coyle K. O. (1990) High productivity of northern Bering Sea benthic amphipods. *Nature* **862**(6269), 862-864.

Hirano H., Komatsu, S., Takakura, H., Sakiyama, F., Tsunasawa, S. (1992) Deblocking and subsequent microsequence analysis of N-alpha-blocked proteins electroblotted onto PVDF membrane. *Journal of Biochemistry* **111**(6), 754-757.

Horiuchi T., Takano, Y., Ishibashi, J., Marumo, K., Kobayashi, K. (2004) Amino acids in water samples from deep sea hydrothermal vents at Suiyo Seamount, Izu-Bonin Arc, Pacific Ocean. *Organic Geochemistry* **35**, 1121-1128.

Horsfall I. M., Wolff, G. A. (1997) Hydrolysable amino acids in sediments from the Porcupine Abyssal Plain, northeast Atlantic Ocean. *Organic Geochemistry* **26**(5-6), 311-320.

Iglesias-Rodriguez M. D., Halloran, P. R., Rickaby, R. E. M., Hall, I. R., Colmenero-Hidalgo, E., Gittins, J. R., Green, D. R. H., Tyrrell, T., Gibbs, S. J., von Dassow, P., Rehm, E., Armbrust, E. V., Boessenkool, K. P. (2008) Phytoplankton calcification in a high-CO2 world. *Science* **320**(5874), 336-340.

Jain E., Bairoch, A., Duvaud, S., Phan, I., Redaschi, N., Suzek, B. E., Martin, M. J., McGarvey, P., Gasteiger, E. (2009) Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics* **10**(136).

Kan J., Hanson T. E., Ginter J. M., Wang K., Chen F. (2005) Metaproteomic analysis of Chesapeake Bay bacterial communities. *Saline Systems* **1**(7), 1-13.

Kang J. H., Lin, C. J., Chen, J., Liu, Q. (2004) Copper induces histone hypoacetylation through directly inhibiting histone acetyltransferase activity. *Chemico-Biological Interactions* **148**(3), 115-123.

Kapuscinski J. (1995) DAPI: a DNA-specific fluorescent probe. *Biotechnic & Histochemistry* **70**(5), 220-233.

Keil R. G., Montlucon, D. B., Prahl, F. G., Hedges, J. I. (1994) Sorptive preservation of labile organic matter in marine sediments. *Nature* **370**(6490), 549-552.

Keil R. G. (1999) Early diagenesis of amino acids in high organic content marine sediments. *Geochemistry of the Earth Surface* **5**, 259-262.

Keller A., Nesvizhskii, A. I., Kolker, E., Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical Chemistry* **74**(20), 5383-5392.

Kim H. Y., Kim, K. (2003) Protein glycation inhibitory and antioxidative activities of some plant extracts in vitro. *Journal of Agricultural and Food Chemistry* **51**(6), 1586-1591.

Knicker H., Hatcher, P.G. (1997) Survival of protein in an organic-rich sediment: possible protection by encapsulation in organic matter. *Naturwissenschaften* **84**(6), 231-234.

Knicker H. (2000) Solid-state 2-D double cross polarization magic angle spinning N-15 C-13 NMR spectroscopy on degraded algal residues. *Organic Geochemistry* **31**, 337-340.

Ko¨pke B., Wilms, R., Engelen, B., Cypionka, H., Sass, H. (2005) Microbial diversity in coastal subsurface sediments: a cultivation approach using various electron acceptors and substrate gradients. *Applied and Environmental Microbiology* **71**(12), 7819-7830.

Kornfeld R., Kornfeld, S. (1985) Assembly of asparagine-linked oligosaccharides. *Annual Review of Biochemistry* **54**, 631-664.

Krogh A., Larsson, B., von Heijne, G., Sonnhammer, E. L. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**(3), 567-580.

Kuleva N. V., Kovalenko, Z. S. (1997) Change in the functional properties of actin by its glycation in vitro. *Biochemistry-Moscow* **62**(10), 1119-1123.

Kuster B., Hunter, A. P., Wheeler, S. F., Dwek, R. A., Harvey, D. J. (1998) Structural determination of N-linked carbohydrates by matrix-assisted laser desorption/ionization mass spectrometry following enzymatic release within sodium dodecyl sulphate polyacrylamide electrophoresis gels: Application to species-specific glycosylation of alpha(1)-acid glycoprotein. *Electrophoresis* **19**(11), 1950-1959.

Laemmli U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lang M., Kroth, P. G. (2001) Diatom fucoxanthin chlorophyll a/c-binding protein (FCP) and land plant light-harvesting proteins use a similar pathway for thylakoid membrane insertion. *Journal of Biological Chemistry* **276**(11), 7985-7991.

Laver W. G. (1964) Structural studies on the protein subunits from three strains of influenza virus. *Journal of Molecular Biology* **9**, 109-124.

Lee C., Cronin, C. (1982) The vertical flux of particulate organic nitrogen in the sea – decomposition of amino-acids in the Peru upwelling area and the equatorial Atlantic. *Journal of Marine Research* **40**(1), 227-251.

Lee C., Wakeham, S. G., Hedges, J. I. (2000) Composition and flux of particulate amino acids and chloropigments in equatorial Pacific seawater and sediments. *Deep Sea Research, Part I, Oceanographic Research Papers* **47**(8), 1535-1568.

Limmer A. W., Wilson, A. T. (1980) Amino-acids in buried paleosols. *Journal of Soil Science* **31**(1), 147-153.

Lipson D. A., Schadt, C. W., Schmidt, S. K. (2002) Changes in soil microbial community structure and function in an alpine dry meadow following spring snow melt. *Microbial Ecology* **43**, 307-314.

Lomas M. W., Moran, S. B. (2011) Evidence for aggregation and export of cyanobacteria and nano-eukaryotes from the Sargasso Sea euphotic zone. *Biogeosciences* **8**(1), 203-216.

Long R. A., Azam, F. (1996) Abundant protein-containing particles in the sea. *Aquatic Microbial Ecology* **10**, 213-221.

Lourenco S. O., Barbarino, E., Marquez, U. M. Lanfer, Aidar, E. (1998) Distribution of intracellular nitrogen in marine microalgae: basis for the calculation of specific nitrogen-to-protein conversion factors. *Journal of Phycology* **34**(5), 798-811.

Lovvorn J. R., Cooper, L. W., Brooks, M. L., De Ruyck, C. C., Bump, J. K., Grebmeier, J. M. (2005) Organic matter pathways to zooplankton and benthos under pack ice in late winter and open water in late summer in the north-central Bering Sea. *Marine Ecology Progress Series* **291**, 135-150.

Loy T. H., Hardy, B.L. (1992) Blood residues analysis of 90,000 year old stone tools from Tabun Cave, Israel. *Antiquity* **66**, 24.

Lubec G., Afjehi-Sadat, L. (2007) Limitations and Pitfalls in Protein Identification by Mass Spectrometry. *Chemical Reviews* **107**, 3568-3584.

Luna G. M., Manini, E., Danovaro, R. (2002) Large fraction of dead and inactive bacteria in coastal marine sediments: Comparison of protocols for determination and ecological significance. *Applied and Environmental Microbiology* **68**(7), 3509-3513.

Maizel J. V. (2000) SDS polyacrylamide gel electrophoresis. *Trends in Biochemical Sciences* **25**(12), 590-592.

Maley F., Trimble, R. B., Tarentino, A. L., Plummer, T. H., Jr. (1989) Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Analytical Biochemistry* **180**(2), 195-204.

Martin P., Lampitt, R. S., Perry, M. J., Sanders, R., Lee, C., D'Asaro, E. (2011) Export and mesopelagic particle flux during a North Atlantic spring diatom bloom. *Deep-Sea Research Part I-Oceanographic Research Papers* **58**(4), 338-349.

Mayer L., Benninger, L., Bock, M., DeMaster, D., Roberts, Q., Martens, C. (2002) Mineral associations and nutritional quality of organic matter in shelf and upper slope sediments off Cape Hatteras, USA: a case of unusually high loadings. *Deep-Sea Research Part II-Topical Studies in Oceanography* **49**(20), 4587-4597.

Mayer L. M. (1994) Surface area control of organic carbon accumulation in continental shelf sediments. *Geochimica et Cosmochimica Acta* **58**(4), 1271-1284.

Mayer L. M. (1999) Extent of coverage of mineral surfaces by organic matter in marine sediments. *Geochimica et Cosmochimica Acta* **63**, 207-215.

McCarthy M., Pratum, T., Hedges, J., Benner, R. (1997) Chemical composition of dissolved organic nitrogen in the ocean. *Nature* **390**(6656), 150-154.

McCarthy M. D., Hedges, J. I., Benner, R. (1998) Major bacterial contribution to marine dissolved organic nitrogen. *Science* **281**, 231-234.

McRoy C. P. (1987) Global maximum of primary production in the North Bering Sea. *E.O.S. Comm.* **68**, 172.

Miki T., Giuggioli, L., Kobayashi, Y., Nagata, T., Levin, S. A. (2009) Vertically structured prokaryotic community can control the efficiency of the biological pump in the oceans. *Theoretical Ecology* **2**(4), 199-216.

Milligan A. J., Morel, F. M. M. . (2002) A proton buffering role for silica in diatoms. *Science* **297**(5588), 1848-1850.

Mintrop L., Duinker, J. C. (1994) Depth profiles of amino-acids in porewater of sediments from the Norwgian - Greenland Sea. *Oceanologica Acta* **17**(6), 621-631.

Moore E. K., Nunn, B. L., Faux, J. F., Goodlett, D. R., Harvey, H. R. Evaluation of electrophoretic protein extraction and database-driven protein identification from marine sediments *Limnology and Oceanography: Methods* **In Review**.

Moore E. K., Nunn, B. L., Goodlett, D. R., Harvey, H. R. Identifying and tracking proteins through the marine water column: insights into the inputs and preservation mechanisms of protein in sediments. *Geochemica et Cosmochimica Acta* **In Review**.

Morris R. M., Nunn, B. L., Frazar, C., Goodlett, D. R., Ting, Y. S., Rocap, G. (2010) Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME Journal* **4**(5), 673-685.

Müller P. J., Suess, E., Ungerer, C. A. (1986) Amino acids and amino sugars of surface particulate and sediment trap material from waters of the Scotia Sea. *Deep-Sea Research Part A-Oceanographic Research Papers* **33**, 819-838.

Nagata T., Kirchman, D. L. (1996) Bacterial degradation of protein adsorbed to model submicron particles in seawater. *Marine Ecology-Progress Series* **132**(1-3), 241-248.

Nagata T., Fukuda, R., Koike, I., Kogure, K., Kirchman, D. L. (1998) Degradation by bacteria of membrane and soluble protein in seawater. *Aquatic Microbial Ecology* **14**(1), 29-37.

Nesvizhskii A. I., Keller, A., Kolker, E., Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Analytical Chemistry* **75**(17), 4646-4658.

Nguyen R. T., Harvey, H. R. (1997) Protein and amino acid cycling during phytoplankton decomposition in oxic and anoxic waters. *Organic Geochemistry* **27**(3-4), 115.

Nguyen R. T., Harvey, H. R. (2001) Protein preservation in marine systems: hydrophobic and other non-covalent associations as major stabilizing forces. *Geochimica et Cosmochimica Acta* **65**, 1467-1480.

Nguyen R. T., Harvey, H. R. (2003) Preservation via macromolecular associations during Botryococcus braunii decay: proteins in the Pula Kerogen. *Organic Geochemistry* **34**, 1391-1403.

Nunn B. L., Shaffer, S. A., Scherl, A., Gallis, B., Wu, M., Miller, S. I., Goodlett, D. R. (2006) Comparison of a *Salmonella typhimuriumproteome* defined by shotgun proteomics directly on an LTQ-FTand by proteome pre-fractionation on an LCQ-DUO. *Briefings in functional genomics and proteomics* **5**(2), 154-168.

Nunn B. L., Keil, R. G. (2006) A comparison of non-hydrolytic methods for extracting amino acids and proteins from coastal marine sediments. *Marine Chemistry* **98**, 31-42.

Nunn B. L., Timperman, A. T. (2007) Marine Proteomics. *Marine Ecology Progress Series* **332**, 281-289.

Nunn B. L., Aker, J. R., Shaffer, S. A., Tsai, Y. H., Strzepek, R. F., Boyd, P. W., Freeman, T. L., Brittnacher, M., Malmstrom, L., Goodlett, D. R. (2009) Deciphering diatom biochemical pathways via whole cell proteomics. *Aquatic Microbial Ecology* **55**(3), 241-253.

Nunn B. L., Ting, Y. S., Malmstrom, L., Tsai, Y. S., Squier, A., Goodlett, D. R., Harvey, H. R. (2010) The path to preservation: Using proteomics to decipher the fate of diatom proteins during microbial degradation. *Limnology and Oceanography* **55**(4), 1790-1804.

Ogunseitan O. A. (1993) Direct extraction of proteins from environmental samples. *Journal of Microbiological Methods* **17**(4), 273-281.

Ostrom P. H., Schall, M., Gandhi, H., Shen, T., Hauschka, P. V., Strahler, J. R., Gage, D. A. (2000) New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry. *Geochemica et Cosmochimica Acta* **64**(6), 1043-1050.

Oudot-Le Secq M. P., Grimwood, J., Shapiro, H., Armbrust, E. V., Bowler, C., Green, B. R. (2007) Chloroplast genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*: comparison with other plastid genomes of the red lineage. *Molecular Genetics and Genomics* **277**, 427-439.

Pantoja S., Lee, C. (1999) Molecular weight distribution of proteinaceous material in Long Island Sound sediments. *Limnology and Oceanography* **44**(5), 1323-1330.

Passow U., Alldredge, A. L. (1994) Distribution, size, and bacterial colonization of transparent exopolymer particles (TEP) in the ocean. *Marine Ecology Progress Series* **113**, 185-198.

Pederson. (2008) Turning a PAGE: the overnight sensation of SDS-polycrylamide gel electrophoresis. *FASEB Journal* **22**(4), 949-953.

Perkins D. N., Pappin, D. J. C., Creasy, D. M., Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**(18), 3551-3567.

Plummer T. H., Elder, J. H., Alexander, S., Phelant, A. W., Tarentino, A. L. (1984) Demonstration of Peptide:*N*-Glycosidase F activity in Endo-b-*N*-acetylglucosaminidase F preparations. *Biological Chemistry* **259**(17), 10700-10704.

Pomeroy L. R., Deibel, D. (1986) Temperature regulation of bacterial activity during the spring bloom in Newfoundland coastal waters. *Science* **233**(4761), 359-361.

Powell M. J., Sutton, J. N., Del Castillo, C. E., Timperman, A. I. (2005) Marine proteomics: generation of sequence tags for dissolved proteins in sea water using tandem mass spectrometry. *Marine Chemistry* **95**(3-4), 183-198.

Prakash A., Piening, B., Whiteaker, J., Zhang, H., Shaffer,S. A., Martin, D., Hohmann, L., Cooke, K., Olson, J. M., Hansen, S., Flory, M. R., Lee, H., Watts, J., Goodlett, D. R., Aebersold, R., Paulovich, A., Schwikowski, B. (2007) Assessing Bias in Experiment Design for Large Scale Mass Spectrometry-based Quantitative Proteomics. *Molecular & Cellular Proteomics* **6**, 1741-1748.

Pruitt K., Tatusova, T., Maglott, D. (2002) The Reference Sequence (RefSeq) Project. In *The NCBI Handbook* (ed. J. McEntyre, Ostell, J.). National Center for Biotechnology Information.

Quince C., Curtis, T. P., Sloan, W. T. (2008) The rational exploration of microbial diversity. *The ISME Journal* **2**, 1997-2006.

Ragueneau O., Schultes, S., Bidle, K., Claquin, P., La Moriceau, B. (2006) Si and C interactions in the world ocean: importance of ecological processes and implications for the role of diatoms in the biological pump. *Global Biogeochemical Cycles* **20**(GB4S02), doi: 10.1029/2006GB002688.

Rashid M. A., Buckley, D. E., Robertson, K. R. (1972) Interactions of a marine humic acid with clay minerals and a natural sediment. *Geoderma* **8**, 11-27.

Reisfeld R. A., Lewis, U. J., Williams, D. E. (1962) Disc electrophoresis of basic proteins and peptides on polyacrylamide gels. *Nature* **195**, 281-283.

Reuss F. F. (1807) *Mem. Soc. Imperiale Naturalists de Moscow* **2**, 327.

Richier S., Fiorini, S., Kerros, M. E., von Dassow, P., Gattuso, J. P. (2011) Response of the calcifying coccolithophore Emiliania huxleyi to low pH/high pCO(2): from physiology to molecular level. *Marine BIology* **158**(3), 551-560.

Rusch D. B., Halpern, A. L., Sutton, G., Heidelberg, K.. B., Williamson, S., Yooseph, S., Wu, D., Eisen, J. A., Hoffman, J. M., Remington, K., Beeson, K., Tran, B., Smith, H., Badon-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J. E., Li, K., Kravitz, S., Heidelberg, J. F., Utterback, T., Rogers, Y. H., Falcon, L. I., Souza, V., Bonilla-Rosso, G., Eguiarte, L. E., Karl, D. M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M. R., Strausberg, R. L., Nealson, K., Friedman, R., Frazier, M., Venter, J. C. (2007) The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biology* **5**(3), e77.

Sambrotto R. N., Niebauer, H. J., Goering, J. J., Iverson, R. L. (1986) Relationships among verticle mixing, nitrate uptake, and phytoplankton growth during spring bloom in the S-E Bering Sea. *Continental Shelf Research* **5**, 161-198.

Scherl A., Shaffer, S. A., Taylor, G. K., Kulasekara, H. D., Miller, S. I., Goodlett, D. R. . (2008) "Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides." *Analytical Chemistry* **80**(4), 1182-1191.

Schulze W. X., Gleixner, G., Kaiser, K., Guggenberger, G., Mann, M., Schulze, E. D. (2005) A proteomic fingerprint of dissolved organic carbon and of soil particles. *Oecologia* **142**, 335-343.

Shapiro A. L., Vin˜uela, E., Maizel, J. V., Jr. . (1967) Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochemical and Biophysical Research Communications* **28**, 815-820.

Sheridan C. C., Lee, C., Wakeham, S. G., Bishop, J. K. B. (2002) Suspended particle organic composition and cycling in surface and midwaters of the equatorial Pacific Ocean. *Deep Sea Research, Part I, Oceanographic Research Papers* **49**, 1983-2008.

Shevchenko A., Wilm, M., Vorm, O., Mann, M. (1996) Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Analytical Chemistry* **68**(5), 850-858.

Shi D. L., Xu, Y., Hopkinson, B. M., Morel, F. M. M. (2010) Effect of Ocean Acidification on Iron Availability to Marine Phytoplankton. *Science* **327**(5966), 676-679.

Smetacek V. S. (1985) Role of sinking in diatom life-history cycles - ecological, evolutionary, and geological significance. *Marine Biology* **84**(3), 239-251.

Spahr C. S., Davis, M. T., McGinley, M. D., Robinson, J. H., Bures, E. J., Beierle, J., Mort, J., Courchesne, P. L., Chen, K., Wahl, R. C., Yu, W., Luethy, R., Patterson, S. D. (2001) Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry I. Profiling an unfractionated tryptic digest. *Proteomics* **1**(1), 93-107.

Springer A. M., McRoy, C. P., Flint, M. V. (1996) The Bering Sea Green Belt: Shelf-edge processes and ecosystem production. *Fisheries Oceanography* **5**(3-4), 205-223.

Steward G. F., Smith, D. C., Azam, F. (1996) Abundance and production of bacteria and viruses in the Bering and Chukchi Seas. *Marine Ecology Progress Series* **131**(1-3), 287-300.

Sukhanova I. N., Semina, H. J., Venttsel, M. V. (1999) Spatial distribution and temporal variability of phytoplankton in the Bering Sea. In *Dynamics of the Bering Sea* (ed. T. R. Loughlin, Ohtani, K.), pp. 193-215. University of Alaska Sea Grant.

Sukhanova I. N., Flint, M. V., Pautova, L. A., Stockwell, D. A., Grebmeier, J. M., Sergeeva, V. M. (2009) Phytoplankton of the western Arctic in the spring and summer of 2002: Structure and seasonal changes. *Deep-Sea Research Part II-Topical Studies in Oceanography* **56**(17), 1223-1236.

Sulochana K. N., Indra, C., Rajesh, M., Srinivasan, V., Ramakrishnan, S. (2001) Beneficial role of amino acids in mitigating cytoskeletal actin glycation and improving F-actin content: In vitro. *Glycoconjugate Journal* **18**(4), 277-282.

Sultana N., Choudhary, M. I., Khan, A. (2009) Protein glycation inhibitory activities of *Lawsonia inermis* and its active principles. *Journal of Enzyme Inhibition and Medicinal Chemistry* **24**(1), 257-261.

Sun Z., Liu, J., Zeng, X., Huangfu, J., Jiang, Y., Wang, M., Chen, F. (2011) Astaxanthin is responsible for antiglycoxidative properties of microalga *Chlorella zofingiensis*. *Food Chemistry* **126**, 1629-1635.

Suzuki S., Kogure, K., Tanoue, E. (1997) Immunochemical detection of dissolved proteins and their source bacteria in marine environments. *Marine Ecology Progress Series* **158**, 1-9.

Tabb D. L., Eng, J. K., Yates, J. R. III. (2000) Proteome Research: Mass Spectrometry, pp. 125. Springer.

Takahashi K., Fujitani, N., Yanada, M. (2002) Long term monitoring of particle fluxes in the Bering Sea and the central subarctic Pacific Ocean, 1990–2000. *Progress in Oceanography* **55**, 95-112.

Talasz H., Wasserer, S., Puschendorf, B. (2002) Nonenzymatic glycation of histones in vitro and in vivo. *Journal of Cellular Biochemistry* **85**(1), 24-34.

Tanner S., Shu, H. J., Frank, A., Wang, L. C., Zandi, E., Mumby, M., Pevzner, P. A., Bafna, V. (2005) InsPecT: Identification of posttransiationally modified peptides from tandem mass spectra. *Analytical Chemistry* **77**(14), 4626-4639.

Tanoue E. (1992) Occurrence and characterization of particulate proteins in the Pacific Ocean. *Deep-Sea Research Part A-Oceanographic Research Papers* **39**(5), 743-761.

Tuross N., Barnes, I., Potts, R. (1996) Protein identification of blood residues on experimental stone tools. *Journal of Archaeological Science* **23**, 289.

Vaque D., Guadayol, O., Peters, F., Felipe, J., Angel-Ripoll, L., Terrado, R., Lovejoy, C., Pedros-Alio, C. (2008) Seasonal changes in planktonic bacterivory rates under the ice-covered coastal Arctic Ocean. *Limnology and Oceanography* **53**(6), 2427-2438.

Wakeham S. G., Lee, C., Farrington, J. W., Gagosian, R. B. (1984) Biogeochemistry of particulate organic-matter in the oceans - results from sediment trap experiments. *Deep-Sea Research Part A-Oceanographic Research Papers* **31**(5), 509-528.

Wakeham S. G., Lee, C., Hedges, J. I., Hernes, P. J., Peterson, M. L. (1997) Molecular indicators of diagenetic status in marine organic matter. *Geochemica et Cosmochimica Acta* **61**(24), 5363-5369.

Wakeham S. G., Lee, C., Peterson, M. L., Liu, Z. F., Szlosek, J., Putnam, I. F., Xue, J. H. (2010) Organic biomarkers in the twilight zone-Time series and settling velocity sediment traps during MedFlux. *Deep-Sea Research Part II-Topical Studies in Oceanography* **56**(18), 1437-1453.

Waldhier M. C., Dettmer, K., Gruber, M. A., Oefner, P. J. (2010) Comparison of derivatization and chromatographic methods for GC-MS analysis of amino acid enantiomers in physiological samples. *Journal of Chromatography B* **878**(15-16), 1103-1112.

Walsh J. J., MCRoy, C. P., Coachman, L. K., Goering, J. J., Nihoul, J. J., Whitledge, T. E., Blackburn, T. H., Parker, P. L., Wirick, C. D., Shuert, P. G., Grebmeier, J. M., Springer, A. M., Tripp, R. D., Hansell, D. A., Djenidi, S., Deleersnijder, E., Henriksen, K., Lund, B. A., Andersen, P., Mullerkarger, F. E., Dean, K. (1989) Carbon and nitrogen cycling within the Bering Chukchi Seas - Source regions for organic-matter effecting AOU demands of the Arctic-Ocean. *Progress in Oceanography* **22**(4), 277-359.

Wang S. L., Moore, J. K. (2011) Incorporating Phaeocystis into a Southern Ocean ecosystem model. *Journal of Geophysical Research-Oceans* **116** C01019.

Wang X. C., Lee, C. (1993) Adsorption and desorption of aliphatic amines, amino acids and acetate by clay minerals and marine sediments. *Marine Chemistry* **44**, 1-23.

Washburn M. P., Wolters, D., Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* **19**, 242-247.

Wasinger V. C., Cordwekkm S, J., Cerpapoljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L., Humpherysmith, I. (1995) Progress with gene-product mapping of mollecutes – mycoplasma genitalium. *Electrophoresis* **16**(7), 1090-1094.

Westermann M., Rheil, E. (2005) Localisation of fucoxanthin chlorophyll a/c-binding polypeptides of the centric diatom Cyclotella cryptica by immuno-electron microscopy. *Protoplasma* **225**(3-4), 217-223.

Yamauchi Y., Ejiri, Y., Tanaka, K. (2002) Glycation by ascorbic acid causes loss of activity of ribulose-1,5-bisphosphate carboxylase/oxygenase and its increased susceptibility to proteases. *Plant and Cell Physiology* **43**(11), 1334-1341.

Yi E. C., Marelli, M., Lee, H., Purvine, S. O., Aebersold, R., Aitchison, J. D., Goodlett, D. R. (2002) Approaching complete peroxisome characterization by gas-phase fractionation. *Electrophoresis* **23**, 3205-3216.

Yooseph S., Sutton G., Rusch D. B., Halpern A. L., Williamson S. J., Remington K., Eisen J. A., Heidelberg K. B., Manning G., Li W., Jaroszewski L., Cieplak P., Miller C. S., Li H., Mashiyama S. T., Joachimiak M. P., van Belle C., Chandonia J. M., Soergel D. A., Zhai Y., Natarajan K., Lee S., Raphael B. J., Bafna V., Friedman R., Brenner S. E., Godzik A., Eisenberg D., Dixon J. E., Taylor S. S., Strausberg R. L., Frazier M., Venter J. C. (2007) The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biology* **5**(3), 432-466.

Zang X., van Heemst, J. D. H., Dria, K. J., Hatcher, P. G. (2000) Encapsulation of protein in humic acid from a histosol as an explanation for the occurrence of organic nitrogen in soil and sediment. *Organic Geochemistry* **31**(7-8), 679-695.

Zang X., Nguyen, R. T., Harvey, H. R., Knicker, H., Hatcher, P. G. (2001) Preservation of proteinaceous material during the degradation of the green alga Botryococcus braunii: A solid-state 2D ^{15}N ^{13}C NMR Spectroscopy Study. *Geochemica et Cosmochimica Acta* **65**(19), 3299-3305.

Zelles L. (1997) Phospholipid fatty acid profiles in selected members of soil microbial communities. *Chemosphere* **35**, 275-294.

Zubkov M. V., Fuchs, B. M., Eilers, H., Burkill, P. H., Amann, R. (1999) Determination of total protein content of bacterial cells by SYPRO staining and flow cytometry. *Applied and Environmental Microbiology* **65**(7), 3251-3257.